

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: August 20, 2004, 19:45:08 ; Search time 779 Seconds
(without alignments)
10126.963 Million cell updates/sec

Title: US-10-055-624B-1
Doctort 1957

Sequence: 1 gtggaattgcccttcaat.....cctgaaccttatcatgagc 1857

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

```
Minimum DB seq length: 0
Maximum DB seq length: 200000000000
```

Maximum DB seq length: 200000000000

Post-processing:	Minimum Match	0%
	Maximum Match	100%

```
Database : N
1: geneseq.29Jan04.*
2: geneseq.1980s.*
3: geneseq.1990s.*
4: geneseq.2000s.*
5: geneseq.2001as.*
6: geneseq.2001bs.*
7: geneseq.2002s.*
8: geneseq.2003as.*
9: geneseq.2003bs.*
10: geneseq.2004s.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1628	87.7	3082	4	AAH18229	Aah18229 Human cDN
2	636.4	34.3	751	4	AAH08535	Aah08535 Human cDN
3	549.2	29.6	1085	5	AAS29609	Aas29609 Human
4	549	29.6	1118	4	AAS33270	DNA encoc
5	549	29.6	1118	5	AAS29716	Human
6	471.8	25.4	813	3	AAFP1794	Aafp1794 Human bre
7	283.6	15.3	2460	6	AAI39645	Human
8	142.6	7.7	344	3	AACT10622	Human
9	126.6	6.8	330	6	ABL79627	Human
10	124	6.7	553	8	ACH34058	Human
11	117.8	6.3	452	4	AAI12455	Human
12	117.8	6.3	452	4	ABAA4163	Human
13	117.8	6.3	452	4	AAI33809	Human
14	117.8	6.3	452	4	ABAA3706	Human
15	117.8	6.3	452	4	ABAA33908	Human
16	117.8	6.3	452	4	AAK57874	Human
17	117.8	6.3	452	4	AAK02432	Human
18	117.8	6.3	452	4	ABSA27457	Human
19	117.8	6.3	452	5	AAI02366	Human
20	117.8	6.3	452	6	ABSO2333	Human
21	101	5.4	492	4	AAI11890	Human
22	101	5.4	492	4	ABAA53513	Human
23	101	5.4	492	4	AAI53138	Human

[illegible]

CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialized methods. AAH03166 to AAH1628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC present invention

XX Sequence 3082 BP; 628 A; 927 C; 916 G; 611 T; 0 U; 0 Other;

Query Match 87.7%; Score 1628; DB 4; Length 3082;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 1631; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY	19	ATGATCCAGATGTCGGAATATCACTGCGACGGGGCTTGAGCTGTGTTCTCCAAACCGC	78
DB	83	ATGATCCAGATGTCGGAATATCACTGCGACGGGGCTTGAGCTGTGTTCTCCAAACCGC	142
QY	79	ACATCCCGGAAGTACAGCTTACGTCGGGGAACGACAGTGCATGCGACGCGGAGGA	138
DB	143	ACATCCCGGAAGTACAGCTTACGTCGGGGAACGACAGTGCATGCGACGCGGAGGA	202
QY	139	TACCGGAACCCCAAGAGAGTGCATGAGCCAGCTGTGTGCTGCTGCAACCAACCA	198
DB	203	TACCGGAACCCCAAGAGAGTGCATGAGCCAGCTGTGTGCTGCTGCAACCAACCA	262
QY	199	CGTGTGAGTGCAGGAGTGCAGGAGCTGCTCAAGTGAATGACACAGGCTTGCTGCTC	258
DB	263	CGTGTGAGTGCAGGAGTGCAGGAGCTGCTCAAGTGAATGACACAGGCTTGCTGCTC	322
QY	259	GCGGAGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	318
DB	323	GCGGAGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	382
QY	319	CAGACCATTAAGTGTGACAGAGTGTGATATCAAGGCCAAGGTGAACCGGGCTTCAAC	378
DB	383	CAGACCATTAAGTGTGACAGAGTGTGATATCAAGGCCAAGGTGAACCGGGCTTCAAC	442
QY	379	TCCAGCATGAGAGTGGGATCCAGGTGCTGCGAGGACCTGTGCTGAGAAAGCATG	438
DB	443	TCCAGCATGAGAGTGGGATCCAGGTGCTGCGAGGACCTGTGCTGAGAAAGCATG	502
QY	439	AATGTGTGCAAGGCTTGGCCACTTGTGTCGCCCGCGAGAGATCAACCAAGTGAAGCTG	498
DB	503	AATGTGTGCAAGGCTTGGCCACTTGTGTCGCCCGCGAGAGATCAACCAAGTGAAGCTG	562
QY	499	AAGGAGATCAAGCGCGCGGACGAGAGAGGAAGATGAGCAAGTGTGCGCGCTGAAGCGC	558
DB	563	AAGGAGATCAAGCGCGCGGACGAGAGAGGAAGATGAGCAAGTGTGCGCGCTGAAGCGC	622
QY	559	CGGCGCATGCGCTTGTCTATGACAGACACATCAAGGACCTCTGCGCACTGCGCCATT	618
DB	623	CGGCGCATGCGCTTGTCTATGACAGACACATCAAGGACCTCTGCGCACTGCGCCATT	682
QY	619	CAGGCGCATGCGCTTGTCTATGACAGACACATCAAGGACCTCTGCGCACTGCGCCATT	678
DB	683	CAGGCGCATGCGCTTGTCTATGACAGACACATCAAGGACCTCTGCGCACTGCGCCATT	742
QY	679	GAGAGTGTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	738
DB	743	GAGAGTGTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	802
QY	739	CAGATCATGCGCTGATGAGATGTGCGCAATTTGAGCGACGCGCTGCGCTGCTGCTGCTG	798

DB	803	CAGATCATGCGCTGATGAGATGTGCGCAATTTGAGCGACGCGCTGCGCTGCTGCTGCTG	862
QY	799	CACCTTACGCTGAAGGCTTGAATGTTCCATTCCGAGGCTCCGAGTGTGGAGC	858
DB	863	CACCTTACGCTGAAGGCTTGAATGTTCCATTCCGAGGCTCCGAGTGTGGAGC	922
QY	859	CGTGTGAGTGCAGAGGCTTGAATGTTCCATTCCGAGGCTCCGAGTGTGGAGC	918
DB	923	CGTGTGAGTGCAGAGGCTTGAATGTTCCATTCCGAGGCTCCGAGTGTGGAGC	982
QY	919	TGCTGAGAGGCTTGAATGTTCCATTCCGAGGCTCCGAGTGTGGAGC	978
DB	983	TGCTGAGAGGCTTGAATGTTCCATTCCGAGGCTCCGAGTGTGGAGC	1042
QY	979	ATGACCTTTGTGTCTGAGAGCGAGATGACAGCCCGAGTGTGCTGCTGAGTTGAGCC	1038
DB	1043	ATGACCTTTGTGTCTGAGAGCGAGATGACAGCCCGAGTGTGCTGCTGAGTTGAGCC	1102
QY	1039	CAGCCCGGCGATGAGAGGCTGAGTACCGAGAGGCGAGTGCAGAAAGATTCGCTG	1098
DB	1103	CAGCCCGGCGATGAGAGGCTGAGTACCGAGAGGCGAGTGCAGAAAGATTCGCTG	1162
QY	1099	GACAGGAAGTACATGCTGCTCTGTAAGCAGACAGAGTGCCTCTGCTGCTGCTGCTG	1158
DB	1163	GACAGGAAGTACATGCTGCTCTGTAAGCAGACAGAGTGCCTCTGCTGCTGCTGCTG	1222
QY	1159	CCTTGAACCAAGGCTTACCTGAGCTCAATTAAGTCTCTCTCTTGAATGCTTGTGCTG	1218
DB	1223	CCTTGAACCAAGGCTTACCTGAGCTCAATTAAGTCTCTCTCTTGAATGCTTGTGCTG	1282
QY	1219	AAGGACAACTGGAGTGTGCTGCTCTGAGATGAGTGCAGAGTGCCTGCTGCTGCTG	1278
DB	1283	AAGGACAACTGGAGTGTGCTGCTCTGAGATGAGTGCAGAGTGCCTGCTGCTGCTG	1342
QY	1279	GACAAGTCT	1338
DB	1343	GACAAGTCT	1402
QY	1339	CTGCTGCT	1398
DB	1403	CTGCTGCT	1462
QY	1399	CTAGTGCAGAGTGAAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1458
DB	1463	CTAGTGCAGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1522
QY	1459	GGTCAACAAAGCCCGAGGACTTGTGATCTTGTGCTGAGGCGGAGGCTTGTGACAT	1518
DB	1523	GGTCAACAAAGCCCGAGGACTTGTGATCTTGTGCTGAGGCGGAGGCTTGTGACAT	1582
QY	1519	GCGGACCCCTTATGTCATGCGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTG	1578
DB	1583	GCGGACCCCTTATGTCATGCGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTG	1642
QY	1579	GAGTACAGAGCGGAGAGAGACCTCTGCTCAAGCTTCTGCTCTGCGCGAGAGGAGGAG	1638
DB	1643	GAGTACAGAGCGGAGAGAGACCTCTGCTCAAGCTTCTGCTCTGCGCGAGAGGAGGAG	1702
QY	1639	CTGACCAAGTGTGCTGCT 1654	
DB	1703	CTGACCAAGTGTGCTGCT 1718	

RESULT 2
AAH08535
AAH08535 standard; cDNA; 751 BP.
AAH08535;
26-JUN-2001 (first entry)
Human cDNA clone (5'-primer) SEQ ID NO:5370.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Human sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-00116126.

23-JUL-1999; 99JP-00248036.

27-AUG-1999; 99JP-00300253.

11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Satou K, Yamamoto J, Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 1, SEQ ID NO 5370; 2537bp + Sequence listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 751 BP; 166 A; 207 C; 235 G; 140 T; 0 U; 3 Other;

Query Match 34.3%; Score 636.4; DB 4; Length 751;

Best Local Similarity 97.8%; Pred. No. 1.5e-136;

Matches 654; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

19 ATGATCCGAATGTGGGAATATCACTGGGAGCGGGCTTGCTGTTTCCAAACCGC 78

83 ATGATCCGAATGTGGGAATATCACTGGGAGCGGGCTTGCTGTTTCCAAACCGC 142

79 ACATCCCGGAATGACCTTACGTGCGGGGAACGACAGTGCATGGCAGACGGGAGGA 138

143 ACATCCCGGAATGACCTTACGTGCGGGGAACGACAGTGCATGGCAGACGGGAGGA 202

139 TACCGGAATCCCGGAGGTGATGATGACCGAGCTGTGCTCTGCTGCACCAACCA 198

203 TACCGGAATCCCGGAGGTGATGATGACCGAGCTGTGCTCTGCTGCACCAACCA 262

199 CGTGTGAGCTGAGCGTGGGAGGCTGCTCAAGTGGATGACACACGGCTTGCTGCTGCC 258

263 CGTGTGAGCTGAGCGCTGGGAGAGCTGCTCAAGTGAATTGACACACGGCTTGCTGCC 322

259 GCGGAGAGGACCGCTGAGCTGCCCTGCTGTCAGACGCTTCCATGATGATCATTTTGGAG 318

323 GCGGAGAGGACCGCTGAGCTGCCCTGCTGTCAGACGCTTCCATGATGATCATTTTGGAG 382

319 CACACCATTAAGTGTGGACAAAGTGTGATATATCAAGGCCAAGGTGAACCGGGCTTCAAC 378

383 CACACCATTAAGTGTGGACAAAGTGTGATATATCAAGGCCAAGGTGAACCGGGCTTCAAC 442

379 TCCAGCATGAGAGGTGGGACATCCAGGTGCTCGGAGGACCTGTGCTTGAAGACAGTGG 438

443 TCCAGCATGAGAGGTGGGACATCCAGGTGCTCGGAGGACCTGTGCTTGAAGACAGTGG 502

439 AATGTGCAAGGCTTGGCCACCTTGTTGGCCCGCGGAGATCAACCAAGTGAACCTG 498

503 AATGTGCAAGGCTTGGCCACCTTGTTGGCCCGCGGAGATCAACCAAGTGAACCTG 562

499 AAGCAGATCAACGCGCGGACAGAAAGAGAGATGAGACAGTGTGCGGCTGAGCGC 558

563 AAGCAGATCAACGCGCGGACAGAAAGAGAGATGAGACAGTGTGCGGCTGAGACGC 622

559 CGGCGCATGCGCTTGTCTATGCAACACCATCAAGACCTCTGCGCAACTGCGCAT 618

623 CGGCGCATGCGCTTGTCTATGCAACACCATCAAGACCTCTGCGCAACTGCGCAT 682

619 CAGGCGCATGCTGG-AGAGCAGAGACTGTAGCCGATGGTGGCGGTGAGAGAACCGTGT 677

683 CAGGCGCATGCTGGAAAGAGAGAGAGACTGTAAACCGCATGTTGCGGCTTAAAGAAACCGTGT 742

678 GAGAGAGTGT 686

743 NGMAATGT 751

RESULT 3

AA529609

ID AA529609 standard; cDNA; 1085 BP.

XX

AC AA529609;

XX

DT 21-NOV-2001 (first entry)

XX

DE Human endocrine polypeptide encoding cDNA SEQ ID No 109.

XX

Endocrine protein; human; mouse; rabbit; goat; horse; food additive; cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotrophic; cerebroprotective; antiproliferative; cytosolic; cardiac; neuroprotective; neuroprotective; neurotrophic; antibacterial; vitucide; fungicide; cancer; ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm; hyperproliferative disorder; breast; liver; cardiovascular disorder; ss; cerebrovascular disorder; nervous system disorder; bacterial infection; fungal infection; viral infection; ocular disorder; endocrine disorder; gastrointestinal disorder; renal disorder; respiratory disorder; wound healing; skin aging; organ transplantation; food preservative; tissue regeneration; anti-fertility.

KW

KW Homo sapiens.

OS

XX

PD WO200155364-A2.

XX

17-JAN-2001; 2001WO-US001308.

XX

31-JAN-2000; 2000US-0179065P.

XX

PR 04-FEB-2000; 2000US-0180628P.

XX

PR 24-FEB-2000; 2000US-0184664P.

XX

PR 02-MAR-2000; 2000US-0186350P.

XX

PR 16-MAR-2000; 2000US-0189874P.

XX

PR 17-MAR-2000; 2000US-0190075P.

XX

PR 18-APR-2000; 2000US-0196123P.

XX

PR 19-MAY-2000; 2000US-0205515P.

mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological condition can be determined by determining the presence or absence of a mutation in an endocrine polypeptide. The treatable disorders include autoimmune diseases such as rheumatoid arthritis, hyperproliferative disorders such as neoplasms of the breast or liver, cardiovascular disorders such as cardiac arrest, cerebrovascular disorders such as cerebral ischaemia, nervous system disorders such as Alzheimer's disease, infections caused by bacteria, viruses and fungi, ocular disorders such as corneal infection, endocrine disorders such as premature labour and infertility, gastrointestinal disorders such as Crohn's disease, renal disorders such as glomerulonephritis and respiratory disorders such as asthma. The polypeptides can also be used to aid wound healing, to prevent skin aging due to sunburn, to maintain organs before transplantation, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at

Query Match	29.6%	Score 549.2;	DB 5;	Length 1085;
Best Local	28.6%	Pred. No. 1.9e-116;		
Matches 554;	Conservative	0;	Mismatches 16;	Indels 0;
				Gaps 0;

[illegible]

XX		Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina,
KW		rheumatoid arthritis; antiarteriosclerotic; cardiac; vascular;
KW		cytoreductive; thrombolytic; antimicrobial; ophthalmological;
KW		cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW		multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW		Gaucher's disease; neurological disease; cerebrovascular disorder;
XX		thrombosis; wound healing; ss.
OS		Homo sapiens.
XX		
PN		WO200155326-A2.
XX		
PD		02-AUG-2001.
XX		
PF		17-JAN-2001; 2001WO-US0001347.
XX		
PR	31-JAN-2000;	2000US-0179065P.
PR	04-FEB-2000;	2000US-0180628P.
PR	24-FEB-2000;	2000US-0184664P.
PR	02-MAR-2000;	2000US-0186350P.
PR	16-MAR-2000;	2000US-0189874P.
PR	17-MAR-2000;	2000US-0190076P.
PR	18-APR-2000;	2000US-0198123P.
PR	19-MAY-2000;	2000US-0205515P.
PR	07-JUN-2000;	2000US-0209467P.
PR	28-JUN-2000;	2000US-0214886P.
PR	30-JUN-2000;	2000US-0215135P.
PR	07-JUL-2000;	2000US-0216647P.
PR	07-JUL-2000;	2000US-0216880P.
PR	11-JUL-2000;	2000US-0217487P.
PR	11-JUL-2000;	2000US-0217496P.
PR	14-JUL-2000;	2000US-0218289P.
PR	26-JUL-2000;	2000US-0220963P.
PR	26-JUL-2000;	2000US-0220964P.
PR	14-AUG-2000;	2000US-0224518P.
PR	14-AUG-2000;	2000US-0224519P.
PR	14-AUG-2000;	2000US-0225213P.
PR	14-AUG-2000;	2000US-0225214P.
PR	14-AUG-2000;	2000US-0225266P.
PR	14-AUG-2000;	2000US-0225287P.
PR	14-AUG-2000;	2000US-0225288P.
PR	14-AUG-2000;	2000US-0225270P.
PR	14-AUG-2000;	2000US-0225447P.
PR	14-AUG-2000;	2000US-0225757P.
PR	14-AUG-2000;	2000US-0225758P.
PR	14-AUG-2000;	2000US-0225759P.
PR	18-AUG-2000;	2000US-0226279P.
PR	22-AUG-2000;	2000US-0226681P.
PR	22-AUG-2000;	2000US-0226688P.
PR	22-AUG-2000;	2000US-0227182P.
PR	23-AUG-2000;	2000US-0227009P.
PR	30-AUG-2000;	2000US-0228924P.
PR	01-SEP-2000;	2000US-0229287P.
PR	06-SEP-2000;	2000US-0229343P.
PR	01-SEP-2000;	2000US-0229344P.
PR	01-SEP-2000;	2000US-0229345P.
PR	05-SEP-2000;	2000US-0229509P.
PR	05-SEP-2000;	2000US-0229513P.
PR	06-SEP-2000;	2000US-0230437P.
PR	06-SEP-2000;	2000US-0230438P.
PR	08-SEP-2000;	2000US-0231432P.
PR	08-SEP-2000;	2000US-0231433P.
PR	08-SEP-2000;	2000US-0231444P.
PR	08-SEP-2000;	2000US-0231413P.
PR	08-SEP-2000;	2000US-0231414P.
PR	08-SEP-2000;	2000US-0232080P.
PR	12-SEP-2000;	2000US-0231688P.
PR	14-SEP-2000;	2000US-0232397P.
PR	14-SEP-2000;	2000US-0232398P.
PR	14-SEP-2000;	2000US-0232399P.
PR	14-SEP-2000;	2000US-0232400P.

```
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236337P.
PR 29-SEP-2000; 2000US-0236357P.
PR 29-SEP-2000; 2000US-0236358P.
PR 29-SEP-2000; 2000US-0236359P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-0239357P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246522P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246538P.
PR 08-NOV-2000; 2000US-0246603P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.

PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
PI WPI; 2001-451931/48.
XX
DR P-PSDB; AAU20561.
XX
PT New nucleic acids and polypeptides, useful for diagnosing, preventing or
XX treating medical conditions.
XX
PS Claim 1; SEQ ID NO 229; 753pp; English.
XX
CC The invention relates to novel isolated nucleic acid molecules (I)
CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC the prevention, treatment and diagnosis of diseases associated with
CC inappropriate expression of secreted proteins. (I) and complementary
CC sequences may also be used as DNA probes in diagnostic assays (e.g.
CC polymerase chain reactions (PCR)) to detect and quantitate the presence
CC of similar nucleic acid sequences (I) in samples, and so which patients may
CC be in need of restorative therapy. (II) may also be used as antigens in
CC the production of antibodies and in assays to identify modulators
CC (agonists and antagonists) of the expression and activity of the secreted
CC proteins. The anti-(II) antibodies and antagonists may also be used to
CC down regulate expression and activity of (II). The anti-(II) antibodies
CC may also be used as diagnostic agents for detecting the presence of (II)
CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
CC disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and
CC thrombosis), infections caused by bacteria, viruses and fungi and ocular
CC disorders (e.g. corneal infections). (I) and (II), agonists, antagonists
CC and antibodies can also be used to promote wound healing, maintain organs
CC before transplantation, and support cell culture of primary tissues.

Query Match 29.6%; Score 549; DB 4; Length 1118;
Best Local Similarity 96.9%; Pred. No. 2,1e-116;
Matches 555; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 1093 CGCCTGACAGAGATATATGTCCTCTGTAAGACAGAGAGTCCCTCGGTGCC 1152
DB 246 CTCCTCCAGAGATATATGTCCTCTGTAAGACAGAGAGTCCCTTCGATCCC 305
QY 1153 TGGACCTTAGACACAGAGTATCTGCTGTAAGACAGAGTCCCTCTGAGATGCTT 1212
DB 306 TGGACCTTAGACACAGAGTATCTGCTGTAAGACAGAGTCCCTCTGAGATGCTT 365
QY 1213 GTGGCAGAGACACTGGGTGCTGCTCGAGATCACTAGCTCCGCTGACCTGTG 1272
DB 366 GTGGCAGAGACACTGGGTGCTGCTCGAGATCACTAGCTCCGCTGACCTGTG 425
QY 1273 GAGATGACAGTTCCTCTCTTCCACTGAGATGAGTGGSCATGGATGACCCGAG 1332
DB 426 GAGATGACAGTTCCTCTCTTCCACTGAGATGAGTGGSCATGATGACAGCAG 485
QY 1333 GCCTTCCTGCTGCTCTCTGACCTGCTGACAGGCGCAGATGGAGACAGACTACCGAGC 1392
DB 486 GCCTTCCTGCTGCTCTCTGACCTGCTGACAGGCGCAGATGGAGACAGACTACCGAGC 545
QY 1393 GTGAGCTTAGTGACAGAGTAGACGAGAGCGCCATCTACCACTCACAGCCTGCTC 1452
```


08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246603P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451936/48.
DR P-PSDB; AAU18487.
XX
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders of the endocrine system such as reproductive disorders,
PT endocrine cancers and also for testing and detection e.g. diagnosis.
XX
XX
XX
PS Claim 1; SEQ ID NO 216; 604pp; English.
XX
XX
CC Sequences AAS29511-AAS29736 represent cDNA molecules, which encode the
CC endogenous polypeptides of the invention. Endocrine polypeptides and their
CC associated polynucleotides of the invention are useful in the diagnosis,
CC treatment and prevention of various types of disorders in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by determining the presence or
CC absence of a mutation in an endocrine polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma. The polypeptides can also be used
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, to regenerate tissues and in chemotaxis.
CC The polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at

Query Match	29.6%	Score 549;	DB 5;	Length 118;
Best Local Similarity	96.9%	Pred. No. 2.1e-116;		
Matches 555; Conservative	4;	Mismatches 14;	Indels 0;	Gaps 0;

QY 1093 CGCCTGGACAGGAAGTACATCGTGTCTCTGTAAGCAGACAGAGGTGCCCTCTCCGTCCC 1152

Db 246 CTCTTCCGAGGATACATCGTGTCTGTAGCAGACAGAGGTGCCCTTTCGTCCTCC 305

[illegible]

1213 GTGGCCAAGCACTGGGTCTGTCTCTCGAGATCAGTCAGGTCGGCCTGTACACTCTG 1272

Db 366 GTGCCAAGCACTGGTCTCTCCTCGAGATCAGTCAGGTCCGCTGTACTCTG 425

QY 1273 GAGGATGACAGTTCCTCTCCCTCCACATGGAGATGGTGTGCATGTGGATGCAGCCAG 1332

Db 426 GAGGATGACAACTTCCTCTCCCTTCCACATGGAGATGGTGGTCATGTGGATGCAGMCCAG 485

[illegible][illegible]

Db 546 GTGGAGCTAGTGCAGCAGGTAGACGAGGACGACCCATCTACCAAGTCAACGAGCCCTGCC 605

1453 CTCGAGGTACACAAAGCCCCCAGGACTTCGTGATCCTGGCCTCGAGGCGGAAGCCTTGT 1512

Db 606 CTCGGAGGT CACACAAAGCCCCAGGACTTCGTGATCCCTGGCCCTCGAGGCCGAAGCCCTTGT 665

1513 GACAAATGGGGACCCCTATGTCATCGCGCTGAGGTCGGTCACGCTGCCCACACACACCGAG 1514

[illegible]

Dh 726 ACCCGAGTACAGACCGGAGAGACCCTCTGCTCAGGCTTCTGCCCTCTGCCCGGAGGG 785

1633 GACCAGCTGACCAAGTCTGCTGGGTTAGGGTC 1665

Db 786 GACCAGCTGACCAAGTAGCCTGTAGTAGACTC 818

RESULT 6

AAFC1794	ID	AAFC1794	standard; DNA; 813 BP.

AC AAF21794;
xy

DT 27-MAR-2001 (first entry)
XX

[illegible]

KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antiulcer; vulnervary; anticonvulsant;
KW

KW Addison's disease; allergy; autoimmune haemolytic anaemia; antibacterial; antitubercal; antiparasitic; caradant; immune disorder;

multiple sclerosis; rheumatoid arthritis; ulcerative colitis;

XX
OS Homo sapiens.

AA	
PN	WO200055173-A1.

PD 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US005881.
 PE
 XX
 PR 12-MAR-1999; 99US-0124270P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-611515/58.
 DR P-PSDB; AAB58891.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention, treatment
 PT and diagnosis of cancer, immune disorders, cardiovascular disorders and
 PT neurological diseases.
 XX
 PS Claim 1; Page 616-617; 1299pp; English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive; nocitropic;
 CC neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
 CC antiinflammatory; antitumor; vulnary; anticonvulsant; antibacterial;
 CC antifungal; antiparasitic and cardiant activity. The polynucleotide and
 CC protein sequences are used in the diagnosis of cancer, particularly
 CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists
 CC and agonists may also be used in the diagnosis, prevention and treatment
 CC of immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC cardiovascular disorders such as myocardial ischemias, wound healing;
 CC neurological diseases such as cerebral anoxia and epilepsy; and
 CC infectious diseases
 XX
 XX Sequence 813 BP; 162 A; 242 C; 251 G; 148 T; 0 U; 10 Other;
 SO
 Query Match 25.4%; Score 471.8; DB 3; Length 813;
 Best Local Similarity 96.4%; Pred. No. 1.1e-96;
 Matches 487; Conservative 5; Mismatches 12; Indels 1; Gaps 1;
 QY 1162 AGCAACAGGAGTGTACCTGAGCTACATTAAGTCTCTCTTGAAGATGTTGTGGCAAG 1221
 DB 9 AGAGACCAAGGTGTCCTGAGCTACATTAAGTCTCTCTTGAAGATGTTGTGGCAAG 68
 QY 1222 GACAACTGGGTGCTGCTCTCGAGAGATGATGAGTCCGCTGTACCTCTGAGAGATGAC 1281
 DB 69 GACAACTGGGTGCTGCTCTCGAGAGATGATGAGTCCGCTGTACCTCTGAGAGATGAC 128
 QY 1282 AAGTTCTCTCTCTCTCCATGAGAGATGATGATGATGATGAGAGCCCTTCTCTG 1341
 DB 129 AAGTTCTCTCTCTCTCCATGAGAGATGATGATGATGATGAGAGCCCTTCTCTG 188
 QY 1342 CTGCTCTCGAGAGCTGCTGAGAGCCGAGAGTGGACAAGACTTCCGAGCGGTGAGACTA 1401
 DB 189 CTGCTCTCGAGAGCTGCTGAGAGCCGAGAGTGGACAAGACTTCCGAGCGGTGAGACTA 248
 QY 1402 GTGACAGAGGTAGACGA-GGACGAGCCATCTTACACAGTACACAGCCCTTCTGAGAG 1460
 DB 249 GTGACAGAGGTAGACGA-GGACGAGCCATCTTACACAGTACACAGCCCTTCTGAGAG 308
 QY 1461 TCACACAAAGCCCGAGAGACTTTCGATCCCTGCTCGAGCGGAGAGCTTGTACATATG 1520
 DB 309 TCACACAAAGCCCGAGAGACTTTCGATCCCTGCTCGAGCGGAGAGCTTGTACATATG 368
 QY 1521 GGACCCCTATGTATCGGCTGAGAGTGTGTCACGCTGCCACACACAGGAGAGCGCCAGA 1580
 DB 369 GGACCCCTATGTATCGGCTGAGAGTGTGTCACGCTGCCACACACAGGAGAGCGCCAGA 428

QY 1581 GTACAGACGGGAGAGACCCCTCTGCTCAGGCTTGTGCTTGTGCGGCAAGGAGCCAGCT 1640
 DB 429 GTACAGACGGGAGAGACCCCTCTGCTCAGGCTTGTGCTTGTGCGGCAAGGAGCCAGCT 488
 QY 1641 GACCAAGTGTCTGCGGAGTGTAGGCTC 1665
 DB 489 GACCAAGTGTCTGAGTGTAGTACTC 513
 RESULT 7
 ID AAL39645/c
 XX AAL39645 standard; DNA; 2460 BP.
 AC AAL39645;
 XX
 DT 05-SEP-2002 (first entry)
 XX
 DE Human secreted protein DNA SEQ ID No 89.
 XX
 KW Antiarteriosclerotic; cytosolic; HIV; antiallergic; antianaemic;
 KW antiaesthetic; cardiant; vasotropic; neuroprotective; nocitropic; SECP;
 KW anticonvulsant; antiparkinsonian; cerebroprotective; antiinflammatory;
 KW immunosuppressive; human secreted protein; cell proliferative disorder;
 KW arteriosclerosis; cancer; autoimmune; inflammatory disorder; AIDS;
 KW allergy; anaemia; asthma; cardiovascular disease; developmental disorder;
 KW ischaemic heart disease; congestive heart failure; neurological disorder;
 KW renal tubular acidosis; hypothyroidism; Alzheimer's disease; dementia;
 KW Parkinson's disease; epilepsy; stroke; knockin humanised animal;
 KW transgenic animal; gene therapy; gene; ds.
 OS
 XX Homo sapiens.
 XX
 PM WO200238602-A2.
 XX
 PD 16-MAY-2002.
 XX
 XX 08-NOV-2001; 2001WO-US047420.
 PF
 XX
 PR 08-NOV-2000; 2000US-0247505P.
 PR 09-NOV-2000; 2000US-0248642P.
 PR 16-NOV-2000; 2000US-0249874P.
 PR 21-NOV-2000; 2000US-0252824P.
 PR 08-DEC-2000; 2000US-0254305P.
 PR 18-DEC-2000; 2000US-0256448P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 XX Yue H, Yao MG, Gandhi AR, Baughn MR, Swarnakar A, Walla NK,
 PI Sanjanwala M, Thornton M, Elliott VS, Lu Y, Gietzen KU, Burford N,
 PI Ding L, Hafalia AJA, Tang YT, Bandman O, Warren BA, Honchell CD,
 PI Lu DM, Thangavelu K, Lee S, Xu Y, Yang J, Lal PG, Tran B,
 PI Ison CH, Duggan BM, Saperstein SK;
 XX
 DR WPI; 2002-519296/55.
 DR P-PSDB; AAO21684.
 XX
 PT Human secreted proteins and polynucleotides for diagnosing, treating or
 PT preventing disorders of cell proliferative, cardiovascular,
 PT developmental, neurological and autoimmune/inflammatory disorders.
 XX
 PS Claim 5; Page 206-207; 229pp; English.
 XX
 CC The invention relates to an isolated human secreted protein (SECP)
 CC polypeptide from 63 fully defined protein sequences given in the
 CC specification. The polypeptide is useful for the diagnosing/treating of a
 CC disease with decreased/overexpression of SECP. Examples of disorders
 CC associated with abnormal expression of SECP include a cell proliferative
 CC disorder e.g. arteriosclerosis, cancer; autoimmune/inflammatory
 CC disorder, AIDS, allergies, anaemia, asthma; cardiovascular disease e.g.
 CC congestive heart failure, ischaemic heart disease; developmental disorder
 CC e.g. renal tubular acidosis, hypothyroidism; neurological disorder e.g.
 CC Alzheimer's disease, dementia, Parkinson's disease, epilepsy or stroke.
 CC The SECP polynucleotide and polypeptide are further useful for analysing

CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX
SQ Sequence 390 BP; 96 A; 100 C; 86 G; 108 T; 0 U; 0 Other;
Query Match 6.8%; Score 126.6; DB 6; Length 390;
Best Local Similarity 87.9%; Pred. No. 2.5e-19;
Matches 138; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 1644 CAGTGTGCTGGGTAGGGTCTCCCTGACTGAGCTGGTCTGGCAAGTGGCTTATTC 1703
Db 357 CTAGCTTCATTGGGTAGAGTCTCCCGAGCGAGCTGGTCTGGCAAGTGGCTTATTC 238
QY 1704 CTGAGGAGCTGCAATCCAGTCCAAAGGGTCCAGAGCGAGGTTGGAATGAAACTAGC 1763
Db 297 GTGAGAGCTGCAATCCAGTCCAAAGGGTCTGGAGCGAGGTTGGAATGAAACTAGC 238
QY 1764 TGGAGAGCACCTGAGTACTCTTAAAGCAATCCCGTG 1800
Db 237 TGGAGAGCACCAAGTACTCTTAAAGCAATCCCGTG 201
RESULT 10
ACH34058
ID ACH34058 standard; cDNA; 543 BP.
XX ACH34058;
AC 13-OCT-2003 (first entry)
XX Human endothelial cell cDNA #2191.
XX DE Human genome mapping; biodiversity; genetic disorder.
XX KW Human: ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX OS Homo sapiens.
XX PN US2003073623-A1.
XX PD 17-APR-2003.
XX PF 30-JUL-2001; 2001US-00918995.
XX PR 30-JUL-2001; 2001US-00918995.
XX PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridisation probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX Claim 1; SEQ ID NO 21270; 44bp; English.
XX PS The invention relates to an isolated polynucleotide comprising any one of
CC

CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC segdata.uspro.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 543 BP; 141 A; 136 C; 155 G; 108 T; 0 U; 3 Other;
Query Match 6.7%; Score 124; DB 8; Length 543;
Best Local Similarity 75.0%; Pred. No. 1.1e-18;
Matches 195; Conservative 0; Mismatches 60; Indels 5; Gaps 3;
QY 1601 TCTGCTCAGGCTTGTGCTTGCGCGGAGGGGACCACTGACCAAGTCTGCTGGGTTA 1660
Db 34 TCGCACAGAGAGCTGGCGCGCTGACGATGATGCTGCTCAGTGCACATCGGGTTA 93
QY 1661 GGGTCTCCCTGACTGAGCTGGTCTGGCAAGTGGCTTATTCCTGGGGGCTGCAATCCA 1720
Db 94 GGGTCTCCACACCGAGCTGGTCTCAGCAAGTGGTCTCAGTAT-GGGCTGCAACTTG 152
QY 1721 GGTCAAGGGTCCAGAGCGAGCGATTGG--AATGAAAACTAGCTGAGAGCACCTGA 1777
Db 153 GGTGAAAGGGTCCAGAGCGAGCGATTGGAAACGTGGAATAGCTGAGAGCACCTGA 212
QY 1778 GTACTCTTAAGCAATCCCCGTGGCCAAATCAACGCCGATTGGATTACCTTCAAGACA 1837
Db 213 GTACTCTT-AAAGCAATCCCCGTGGACCAATCAACTGCAATTGGATTATCATCGAGCA 271
QY 1838 CCTGAACCTTATCATGAGC 1857
Db 272 CCTGAACCTTATCATGAGC 291
RESULT 11
AA112455
ID AA112455 standard; DNA; 452 BP.
XX AA112455;
AC 12-OCT-2001 (first entry)
XX DE Probe #2388 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX OS Homo sapiens.
XX PN MO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000670.
XX PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-0063236P.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX

(MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX

PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX

PS Claim 25; SEQ ID NO 2388; 487bp; English.

CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 452 BP; 115 A; 88 C; 123 G; 126 T; 0 U; 0 Other;

Query Match 6.3%; Score 117.8; DB 4; Length 452;
Best Local Similarity 81.8%; Pred. No. 2.8e-17;
Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 1644 CAAAGTGTGCTGGGTTAGGCTCTCCCTGACCTGACCTGCTCGGCAAGTGGCTTCTATTTC 1703
DB 70 CTAATGTCACCTGGGTTAGGCTCTCCCTGACCTGCTCGGCAAGTGGCTTCTATTTC-TC 128

QY 1704 CTGGGGGCTCGAATCCAGTCAAGGTCGACGAGGACGCGTTGGAATGAAACTAGC 1763
DB 129 GTGGGGGCTCGAATCCAGTCAAGTCTTCTGAGCGATGTTGGAATGAAACTAGC 188

QY 1764 TGAGGACACCTGAGTACTCTTAAGCAATCCCGTGCCAAATTCACAGCCGATTGG 1823
DB 189 TGAGGACACCGGAGTACTCTTAAGCAATCTCTGTGTGAGTAAGAGGGGAGCTAGGA 248

QY 1824 A 1824
DB 249 A 249

RESULT 12
ABA54163
ID ABA54163 standard; DNA; 452 BP.
XX
AC ABA54163;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #2468.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0068408.
PR 03-AUG-2000; 2000US-0063236P.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX

(MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX

PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX

PS Claim 1; SEQ ID NO 2468; 639pp + Sequence Listing; English.

CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 452 BP; 115 A; 88 C; 123 G; 126 T; 0 U; 0 Other;

Query Match 6.3%; Score 117.8; DB 4; Length 452;
Best Local Similarity 81.8%; Pred. No. 2.8e-17;
Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 1644 CAAATGCTGCTGGGTTAGGCTCTCCCTGACCTGACCTGCTCGGCAAGTGGCTTCTATTTC 1703
DB 70 CTAATGTCACCTGGGTTAGGCTCTCCCTGACCTGCTCGGCAAGTGGCTTCTATTTC-TC 128

QY 1704 CTGGGGGCTCGAATCCAGTCAAGGTCGACGAGGACGCGTTGGAATGAAACTAGC 1763
DB 129 GTGGGGGCTCGAATCCAGTCAAGTCTTCTGAGCGATGTTGGAATGAAACTAGC 188

QY 1764 TGAGGACACCTGAGTACTCTTAAGCAATCCCGTGCCAAATTCACAGCCGATTGG 1823
DB 189 TGAGGACACCGGAGTACTCTTAAGCAATCTCTGTGTGAGTAAGAGGGGAGCTAGGA 248

QY 1824 A 1824
DB 249 A 249

RESULT 13
AAI3809
ID AAI3809 standard; DNA; 452 BP.
XX
AC AAI3809;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #2495 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0068408.
PR 03-AUG-2000; 2000US-0063236P.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX

(MOLE-) MOLECULAR DYNAMICS INC.

```

XX  Penn SG, Hanzel DK, Chen W, Rank DR;
PT  WPI; 2001-488897/53.
XX
XX
PT  Human genome-derived single exon nucleic acid probes useful for analyzing
PT  gene expression in human placenta.
XX
XX  Claim 25; SEQ ID NO 2495; 654bp; English.
XX
CC  The present invention relates to single exon nucleic acid probes (SENP).
CC  The present sequence is one such probe. The probes are useful for
CC  producing a microarray for predicting, measuring and displaying gene
CC  expression in samples derived from human placenta. The probes are useful
CC  for antenatal diagnosis of human genetic disorders
XX
SQ  Sequence 452 BP; 115 A; 88 C; 123 G; 126 T; 0 U; 0 Other;

Query Match      6.3%; Score 117.8; DB 4; Length 452;
Best Local Similarity 81.8%; Pred. No. 2.8e-17;
Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY  1644 CAAGTGTGCTGGGTTAGGGTCTCCCTGACTGAGCTGTGCTCGGCAAGTGGCTTCTATTC 1703
    |||||
DB  70 CTAGTGCACCTGGGTTAGGGTCTCCCTGACCGAGCTGTCTCGCAAGTGTGTCCA-TC 128

QY  1704 CTGGGGGCTCGAATCCAGTCAAGGCTCCGAGAGCGATGGAATGAAAACCTAGC 1763
    |||||
DB  129 GTGGGGGCTCGAATCCAGATCCAGATCGAAGTCTTTCTGAGCGATGTGGAATGAAAACCTAGC 188

QY  1764 TGGAGGACACCTGAGTACTCTTAAAGCAATCCCGTGGCCAAATCAACAGCCGATTGG 1823
    |||||
DB  189 TGGAGGACACCCGAGTACTCTTAAAGCAATCTGTGTGATGAAGAGGAGCTAGGA 248

QY  1824 A 1824
    |
DB  249 A 249

RESULT 14
ABAA3706
ID  ABA43706 standard; DNA; 452 BP.
XX
XX  ABA43706;
XX
XX  01-FEB-2002 (first entry)
XX
DE  Human breast cell single exon nucleic acid probe #2401.
XX
XX  Human; microarray; single exon probe; gene expression; breast; disease;
XX  cancer; ss.
XX
XX  Homo sapiens.
XX
XX  WO200157271-A2.
XX
XX  09-AUG-2001.
XX
XX  30-JAN-2001; 2001WO-US000662.
XX
XX  04-FEB-2000; 2000US-0180312P.
XX  26-MAY-2000; 2000US-0207456P.
XX  30-JUN-2000; 2000US-00608408.
XX  03-AUG-2000; 2000US-00632366.
XX  21-SEP-2000; 2000US-0234687P.
XX  27-SEP-2000; 2000US-0236359P.
XX  04-OCT-2000; 2000GB-00024263.
XX
XX  (MOLB-) MOLECULAR DYNAMICS INC.
XX
XX  Penn SG, Hanzel DK, Chen W, Rank DR;
XX  WPI; 2001-496933/54.

```

```

XX  New spatially-addressable set of single exon nucleic acid probes, useful
PT  for measuring gene expression in sample derived from human breast,
PT  comprises number of single exon nucleic acid probes.
XX
XX
PS  Claim 1; SEQ ID NO 2401; 327bp + Sequence Listing; English.
XX
XX  The invention relates to a spatially-addressable set of single exon
CC  nucleic acid probes for measuring gene expression in a sample derived
CC  from human breast and BT 474 cells. The method involves contacting the
CC  probes with a collection of detectably labelled nucleic acids derived
CC  from mRNA of human breast, and then measuring the label bound to each
CC  probe of the microarray. The probes are useful for verifying the
CC  expression of regions of genomic DNA predicted to encode proteins. They
CC  are useful for gene discovery, and for determining predisposition and/or
CC  prognosing breast disease. Gene expression analysis is useful for
CC  assessing the toxicity of chemical agents on cells. The microarray of
CC  this invention presents a far greater diversity of probes for measuring
CC  gene expression, with far less bias than expressed sequence tag
CC  microarrays. The method is suitable for rapid production of functional
CC  information from genomic sequence. The present sequence is a single exon
CC  nucleic acid probe of the invention. Note: The sequence data for this
CC  patent did not form part of the printed specification, but was obtained
CC  in electronic format directly from WIPO at
    ftp.wipo.int/pub/published_pct_sequences
XX
SQ  Sequence 452 BP; 115 A; 88 C; 123 G; 126 T; 0 U; 0 Other;

Query Match      6.3%; Score 117.8; DB 4; Length 452;
Best Local Similarity 81.8%; Pred. No. 2.8e-17;
Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY  1644 CAAGTGTGCTGGGTTAGGGTCTCCCTGACTGAGCTGTGCTCGGCAAGTGGCTTCTATTC 1703
    |||||
DB  70 CTAGTGCACCTGGGTTAGGGTCTCCCTGACCGAGCTGTCTCGCAAGTGTGTCCA-TC 128

QY  1704 CTGGGGGCTCGAATCCAGTCAAGGCTCCGAGAGCGATGGAATGAAAACCTAGC 1763
    |||||
DB  129 GTGGGGGCTCGAATCCAGATCCAGATCGAAGTCTTTCTGAGCGATGTGGAATGAAAACCTAGC 188

QY  1764 TGGAGGACACCTGAGTACTCTTAAAGCAATCCCGTGGCCAAATCAACAGCCGATTGG 1823
    |||||
DB  189 TGGAGGACACCCGAGTACTCTTAAAGCAATCTGTGTGATGAAGAGGAGCTAGGA 248

QY  1824 A 1824
    |
DB  249 A 249

RESULT 15
ABA23908
ID  ABA23908 standard; DNA; 452 BP.
XX
XX  ABA23908;
XX
XX  23-JAN-2002 (first entry)
XX
DE  Probe #2374 for gene expression analysis in human heart cell sample.
XX
XX  Human; gene expression; heart; microarray; vascular system; probe;
XX  cardiovascular disease; hypertension; cardiac arrhythmia;
XX  congenital heart disease; ss.
XX
XX  Homo sapiens.
XX
XX  WO200157274-A2.
XX
XX  09-AUG-2001.
XX
XX  30-JAN-2001; 2001WO-US000666.
XX
XX  04-FEB-2000; 2000US-0180312P.
XX  26-MAY-2000; 2000US-0207456P.

```

PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOIE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human hearts.

PS Claim 1; SEQ ID NO 2374; 530bp; English.

XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 452 BP; 115 A; 88 C; 123 G; 126 T; 0 U; 0 Other;

Query Match 6.3%; Score 117.8; DB 4; Length 452;

Best Local Similarity 81.8%; Pred. No. 2.8e-17;

Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 1644 CAAGTGTGCTGGGTTAGGGTCTCCCTGACTGAGCTGGTCTGGCAAGTGGCTTCTATTG 1703

Db 70 CTAGTGCACCTGGGTTAGGGTCTCCCTGACGAGCTGGTCTGGCAAGTGTGTCCA-TC 128

QY 1704 CTGGGGGGCTCGAATCGAGTCAAGGGTCGACGAGCGACGGTTGGAAATGAAAACTAGC 1763

Db 129 GTGGGGGGCTCGAATCGAATCGAAGTGTCTTGGAGCGATGGTGGAAATGAAAACTAGC 188

QY 1764 TGGAGGACACCTGAGTACTCTTAAAGCAATCCCGTGGCCAAATCAACAGCCGATTGG 1823

Db 189 TGGAGGACACCGAGTACTCTTAAAGCAATCTCTGTGAGTAAAGGAGGAGCTAGGA 248

QY 1824 A 1824

Db 249 A 249

Search completed: August 20, 2004, 21:31:04
 Job time : 787 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2004, 19:45:09 ; Search time 7436 Seconds

(without alignments)
10824.101 Million cell updates/sec

Title: US-10-055-624B-1

Perfect score: 1857
Sequence: 1 gtgaattgccttcaaat.....cctgaacctatcatgagc 1857

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pac: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1857	100.0	1857	AF416921	AF416921 Homo sapi
2	1839	99.0	6359	AB014607	AB014607 Homo sapi
3	1633.6	88.0	1818	AF416922	AF416922 Homo sapi
4	1628	87.7	3082	AX883384	AX883384 Sequence
5	1628	87.7	3082	BD160291	BD160291 Primer fo
6	1628	87.7	3082	AK023937	AK023937 Homo sapi
7	1450	78.1	2875	AK091112	AK091112 Homo sapi
8	1230	66.2	2678	AF416923	AF416923 Mus muscu
9	1230	66.2	5606	BC042492	BC042492 Mus muscu
10	1182.2	63.7	5415	AK122350	AK122350 Mus muscu
11	730.8	39.4	2455	BC001517	BC001517 Homo sapi
12	636.4	34.3	751	AX870465	AX870465 Sequence
13	636.4	34.3	751	BD150527	BD150527 Primer fo
14	434.8	23.4	1689	AB078618	AB078618 Mus muscu
15	433.6	23.3	1725	AB040609	AB040609 Rattus no
16	425	22.9	3019	BC025852	BC025852 Mus muscu
17	405	21.8	1820	AB078619	AB078619 Homo sapi
18	405	21.8	2077	AK122960	AK122960 Homo sapi
19	211.4	11.4	159681	AC036239	AC036239 Homo sapi
20	209.8	11.3	122146	AC011736	AC011736 Homo sapi
21	209.8	11.3	132033	AC099796	AC099796 Homo sapi
22	158.4	8.5	62485	AL590093	AL590093 Human DNA
23	156.8	8.4	159681	AC036239	AC036239 Homo sapi
24	143.2	7.7	2242	AX834086	AX834086 Sequence
25	143.2	7.7	2242	AK096315	AK096315 Homo sapi
26	143.2	7.7	246011	AL929585	AL929585 Mouse DNA
27	142.6	7.7	344	AX898834	AX898834 Sequence
28	142.6	7.7	344	BD034367	BD034367 Rattus no
29	140.8	7.6	252616	AC122593	AC122593 Rattus no
30	134.6	7.2	169770	AC021763	AC021763 Homo sapi
31	134.6	7.2	174741	AC040977	AC040977 Homo sapi
32	132.6	7.1	162914	AC008813	AC008813 Homo sapi
33	131.4	7.1	77001	AL513013	AL513013 Human DNA
34	131.4	7.1	154361	AC092061	AC092061 Homo sapi
35	131.4	7.1	167331	AC073115	AC073115 Homo sapi
36	131.4	7.1	177492	AC073350	AC073350 Homo sapi
37	130.6	7.0	147181	AL157931	AL157931 Human DNA
38	130.4	7.0	172274	AP000924	AP000924 Homo sapi
39	130.4	7.0	174084	AC084395	AC084395 Homo sapi
40	130.4	7.0	174311	AP002788	AP002788 Homo sapi
41	130.4	7.0	190592	AC015806	AC015806 Homo sapi
42	129.8	6.9	170470	AC092364	AC092364 Homo sapi
43	128.2	6.9	180510	AC022415	AC022415 Homo sapi
44	125.8	6.8	92941	AC011313	AC011313 Homo sapi
45	125	6.7	189278	AC055787	AC055787 Homo sapi

ALIGNMENTS

RESULT 1
AF416921
LOCUS AF416921 1857 bp mRNA linear PRI 13-DEC-2001
DEFINITION Homo sapiens brown fat inducible thioesterase 1 mRNA, complete cds,
alternatively spliced.
ACCESSION AF416921
VERSION AF416921.1 GI:17646236
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1857) Adams,S.H., Chui,C., Schilbach,S.L., Yu,X.X., Goddard,A.D.,
Grimaldi,C., Lee,J., Dowd,P., Colman,S. and Lewin,D.A.

Db 1561 ACAACCGAGAGACGCCAGAGTACAGACGCCGAGAGACCTCTGCTCAGGCTTCTGCTC 1620

QY 1621 TGCGCGAGGGGAGACCACTGACCACTGCTGCTGAGTGGCTCTCCCTGACTGAGCTG 1680

Db 1621 TGCGCGAGGGGAGACCACTGACCACTGCTGCTGAGTGGCTCTCCCTGACTGAGCTG 1680

QY 1681 GTCTCGGCAAGTGGCTTCTATCTCTGAGGAGCTGAACTCCAGGCTCAAGGGTCCAGAGAGC 1740

Db 1681 GTCTCGGCAAGTGGCTTCTATCTCTGAGGAGCTGAACTCCAGGCTCAAGGGTCCAGAGAGC 1740

QY 1741 GACGGTTGGAATGAGAACTAGCTGAGAGACACCTGAGTACTCTTAAAGCAATCCCGGTG 1800

Db 1741 GACGGTTGGAATGAGAACTAGCTGAGAGACACCTGAGTACTCTTAAAGCAATCCCGGTG 1800

QY 1801 GCCAAATCAACGCCGATTTGGATACCTTCAAGACACCTGAACTTATCATGAGC 1857

Db 1801 GCCAAATCAACGCCGATTTGGATACCTTCAAGACACCTGAACTTATCATGAGC 1857

RESULT 2

AB014607 6359 bp mRNA linear PRI 06-FEB-1999

LOCUS Homo sapiens mRNA for KIAA0707 protein, partial cds.

DEFINITION AB014607

ACCESSION AB014607.1 GI:33272227

VERSION

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Ishikawa, K., Nagase, T., Suyama, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.

TITLE Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro

JOURNAL DNA Res. 5 (3), 169-176 (1998)

MEDLINE 9734811

PUBMED 98403880

REFERENCE 2 (bases 1 to 6359)

AUTHORS Ohara, O., Suyama, M., Nagase, T. and Ishikawa, K.

TITLE Direct Submission

JOURNAL Submitted (26-May-1998) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)

FEATURES

source location/Qualifiers

1..6359

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="HH00286"

/sex="male"

/tissue_type="brain"

/clone_id="pBluescriptII SK plus"

/dev_stage="adult"

1..6359

/gene="KIAA0707"

<1..1895

/gene="KIAA0707"

/note="HH00286 cDNA clone for KIAA0707 has a 1-bp deletion at 1560 of the sequence of KIAA0707."

/codon_start=3

/product="KIAA0707 protein"

/protein_id="BAA31682.1"

/db_xref="GI:33272228"

/translation="PGCCSGDQVLSIGRCFPPPPGAMTQNVGNIHRLGLASVFSNRTSRKSLRAGNDSAMADGEGYRNPTEVQMSQVLPCHTNDGELSVQGLKWDITTAELKSRHAGCPCTVATMDIYFEHTISVGOVNIKAKVAFNSMEWGIQVASEDLGSEKQNNVKALATFVAREITVKLQIIPREBEKMEHSVAERRMILVYADTIKLLASCAIIGDLESRDSCRMVPAEKTRVSEVLVPPHANHGNTFAGGQIMAMENVAITLASFCAHPTKALIMFHPRGSDQVDRVLTKALUNNAFHGSHVEGVCVAYVQBAETHRHINSAPFTFVLDADDPQLPWIROPDGERRYRARSARKKIRIDRKITIVGCK

ORIGIN

Query Match 99.0%; Score 1839; DB 9; Length 6359;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1839; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 ATGATCCAGAAATTCGGAAATCACTGCGACCGGGGCTTGGCTCTGTGTTCTCCAAACCG 78

Db 72 ATGATCCAGAAATTCGGAAATCACTGCGACCGGGGCTTGGCTCTGTGTTCTCCAAACCG 131

QY 79 ACATCCCGGAAGTCAGACCTTACGTCGCGGGAGACGACAGTGCCTGAGCAGCGGAGGGA 138

Db 132 ACATCCCGGAAGTCAGACCTTACGTCGCGGGAGACGACAGTGCCTGAGCAGCGGAGGGA 191

QY 139 TACCGGAACCCCAACGAGGTGACAGATGAGCCAGCTGGTGTGCTGCTCCACACCAACCA 198

Db 192 TACCGGAACCCCAACGAGGTGACAGATGAGCCAGCTGGTGTGCTGCTCCACACCAACCA 251

QY 199 CGTGTGAGCTGAGCGCTCGGAGAGCTGCTTAAGTGAATTACACACGCTTGTGCTCC 258

Db 252 CGTGTGAGCTGAGCGCTCGGAGAGCTGCTTAAGTGAATTACACACGCTTGTGCTCC 311

QY 259 GCGGAGAGGACGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318

Db 312 GCGGAGAGGACGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 371

QY 319 CACACCATTAATGTTGAGCAAGTGTGATATATCAAGGCCAAGGTGAACCGGGCTTCAAC 378

Db 372 CACACCATTAATGTTGAGCAAGTGTGATATATCAAGGCCAAGGTGAACCGGGCTTCAAC 431

QY 379 TCCAGATGAGAGTGGGCAATCCAGTGTGCTGCGAGAGACCTGTGCTTGTGAAGCAAGTG 438

Db 432 TCCAGATGAGAGTGGGCAATCCAGTGTGCTGCGAGAGACCTGTGCTTGTGAAGCAAGTG 491

QY 439 AATGTGCAAGGCTTGTGCGCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 498

Db 492 AATGTGCAAGGCTTGTGCGCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 551

QY 499 AAGCAGATCAAGCGCGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 558

Db 552 AAGCAGATCAAGCGCGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 611

QY 559 CGGCGCATGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 618

Db 612 CGGCGCATGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 671

QY 619 CAGGCGCATGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 678

Db 672 CAGGCGCATGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 731

QY 679 GAGAGTGTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738

Db 732 GAGAGTGTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 791

QY 733 CAGATCATGCGCTGAGTGAAGATGTGGCCACCATTCAGACCGGCGCTGCGCGGCG 798

Db 792 CAGATCATGCGCTGAGTGAAGATGTGGCCACCATTCAGACCGGCGCTGCGCGGCG 851

QY 799 CACCTTAAGTGAAGGACCATTTGAATGTTCACCTTCGAGAGGCGGCTCCAGAGTGGAGAC 858

Db 852 CACCTTAAGTGAAGGACCATTTGAATGTTCACCTTCGAGAGGCGGCTCCAGAGTGGAGAC 911

QY 859 CGTCTGAGTCTCAAGCGCATGTGAACATGCTTCAACATATGACATGAGTGGAGTGGCTG 918

Db 912 CGTCTGAGTCTCAAGCGCATGTGAACATGCTTCAACATATGAGTGGAGTGGAGTGGCTG 971

QY 919 TGGTGAAGGCTTATGCGCAGAGGCTGAGACCCACCGGCGCAATCAACAGTGGCTTT 978

Db	972	TGCGTGGAGGCGCTATCGCCAGAGGCTGAGACCAACGGCGGCGACATCAACAGTGCCTTT	1031
QY	979	ATGACCTTTTGGTGTCTTGAGACGCAATGACACAGCCCACTGTGTGCTCCTGGATTGCGCCC	1038
Db	1032	ATGACCTTTTGGTGTCTTGAGACGCAAGTATCCAGCCCACTGTGTGCTCCTGGATTGCGCCC	1091
QY	1039	CAGCCCGGCGCATGGTGAAGCGCGGATCCGAGAGGCGAATGGCGAAGAAACAATCGGCTG	1098
Db	1092	CAGCCCGGCGCATGGTGAAGCGCGGATCCGAGAGGCGAATGGCGAAGAAACAATCGGCTG	1151
QY	1099	GACAGGAAGTACATCGTGTCTCTGTAAAGCAGAGAGTGCCTCTCCGTGCCCTGGAGAC	1158
Db	1152	GACAGGAAGTACATCGTGTCTCTGTAAAGCAGAGAGTGCCTCTCCGTGCCCTGGAGAC	1211
QY	1159	CTTGTGCAACCAAGGTGTACTGTAGTACATATACGTCTCTCTCTTGAAGATGCTGTGGCC	1218
Db	1212	CTTGTGCAACCAAGGTGTACTGTAGTACATATACGTCTCTCTCTTGAAGATGCTGTGGCC	1271
QY	1219	AAGGACAACCTGGTGTCTCTCGGAGATCAGTACAGTCCGCGCTGTACATCTGGAGAT	1278
Db	1272	AAGGACAACCTGGTGTCTCTCGGAGATCAGTACAGTCCGCGCTGTACATCTGGAGAT	1331
QY	1279	GACAAAGTTCCTCTCTCTTCCACATGGAGATGGTGGTGCATGTGTGATGTGACCCAGGCTTC	1338
Db	1332	GACAAAGTTCCTCTCTCTTCCACATGGAGATGGTGGTGCATGTGTGATGTGACCCAGGCTTC	1391
QY	1339	CTGCTGTCTCTCGACCTGCGTCAAGGCGCAAGTGTGGGCAACAGCACTACCGAGCGTGGAG	1398
Db	1392	CTGCTGTCTCTCGACCTGCGTCAAGGCGCAAGTGTGGGCAACAGCACTACCGAGCGTGGAG	1451
QY	1399	CTACTGTGACGAGTGAACGAGAGAGACGCGCATCTAACCGATCAACAGCCCTGGCCCTGGGA	1458
Db	1452	CTACTGTGACGAGTGAACGAGAGAGACGCGCATCTAACCGATCAACAGCCCTGGCCCTGGGA	1511
QY	1459	GGTCAACAACAAAGCCCCAGAGACTTGCTGTACTCTGGGCTGAGAGCGGAACCTTGTGACAT	1518
Db	1512	GGTCAACAACAAAGCCCCAGAGACTTGCTGTACTCTGGGCTGAGAGCGGAACCTTGTGACAT	1571
QY	1519	GGGAGCCCTTATGTTCATCGCGCTGAGGTGCGTCAACGCTGACCAACAACGAGAGACGCCA	1578
Db	1572	GGGAGCCCTTATGTTCATCGCGCTGAGGTGCGTCAACGCTGACCAACAACGAGAGACGCCA	1631
QY	1579	GAGTACACAGCGCGGAGAGACCTCTGTCTCAGGCTTCTGCTCTTGCGCGGAGGGGGACGAG	1638
Db	1632	GAGTACACAGCGCGGAGAGACCTCTGTCTCAGGCTTCTGCTCTTGCGCGGAGGGGGACGAG	1691
QY	1639	CTGACCAAGTCTCTCTGTGGATTAAAGGCTCCCTGACCTGACGTGCTGTGGCAAGTGGCTTC	1698
Db	1692	CTGACCAAGTCTCTCTGTGGATTAAAGGCTCCCTGACCTGACGTGCTGTGGCAAGTGGCTTC	1751
QY	1699	TATTCCTGTGGGGCTCGAATCCAGGTCAAAAGGATCGCAGAGACGACGCTGTGGATGGAATA	1758
Db	1752	TATTCCTGTGGGGCTCGAATCCAGGTCAAAAGGATCGCAGAGACGACGCTGTGGATGGAATA	1811
QY	1759	CTAGCTGAGGAGACCTGAGTACTTTAAAGCAATCCCGTGGGCCAAATCAACGCCGA	1818
Db	1812	CTAGCTGAGGAGACCTGAGTACTTTAAAGCAATCCCGTGGGCCAAATCAACGCCGA	1871
QY	1819	TTTGGATTACCTTCAAGACACCTGTAAACCTTATCACTGAGC	1857
Db	1872	TTTGGATTACCTTCAAGACACCTGTAAACCTTATCACTGAGC	1910

RESULT 3	AF416922	LOCUS	AF416922	1818 bp	mRNA	linear	PRI 13-DEC-2001
DEFINITION			Homo sapiens brown fat inducible thioesterase 2 mRNA, complete cds, alternatively spliced.				

ACCESSION	AF416922
VERSION	AF416922.1
KEYWORDS	GI:17646238
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumaleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (baes 1 to 1818)
AUTHORS Adams,S.H., Chui,C., Schibach,S.L., Yu,X.X., Goddard,A.D., Grimaldi,C., Lee,J., Dowd,P., Colman,S. and Lewin,D.A.
TITLE BP1r, a unique ACyl-CoA thioesterase Induced in Thermogenic Brown Adipose tissue. Cloning, Organization of the Human Gene, & Assessment of a Potential Link to Obesity
JOURNAL Biochem. J. (2002) in press
REFERENCE 2 (baes 1 to 1818)
AUTHORS Adams,S.H., Chui,C., Schibach,S.L., Yu,X.X., Goddard,A.D., Grimaldi,C., Lee,J., Dowd,P., Colman,S. and Lewin,D.A.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2001) Endocrinology, Genentech, Inc., 1 DNA way, South San Francisco, CA 94080, USA
FEATURES location/Qualifiers
source 1..1818

CDS

```

/feature="alternatively spliced"
/codon_start=1
/product="brown fat inducible thioesterase 2"
/protein_id="AAL40938.1"

```

/transal:ia "m1QvNGvNlRbRlAaJvFvSvRtSRvKvLRAvGvNvSvAMvDvGvYvNvPv
 vTevOMvOvLvPvCvHvTlNGvNvSvGvOvLlKvMvDvTlCvLsERvAGvOvCvIvAvmDvDvIvFvEvl
 lsvGvOvGvNvKvKvNvKvNvKvNvSvMvRvMvYvIvADvDvCvSEvKvOMvCvKvIvAvtvARvEvlTtKvL
 KvOvTtPvTtBvEvlKvHsvAARvRvMvRvIvYvADvDvIvKvOvLlNvCvAvtvOvGvLsvEsvDsvRvMvPvAvt
 RvSvSvEvlYvIvPvAHHvOvGvNvGvOvCvOvIvMvMvNvAvtvIvAvtSRvAvtPvIvtvKvIvEMvPvRvSv
 QvUvDvPvIvtvKvAvtYvNvAvtFvRvHsvEsvGvCvAvtOvAvtPvRvHvRvNvSvMvPvIvtvKvIvADvDvOvL
 LvPvIvtvRvQvDvGvErvRvTtEAvRvKvIvRvIvDvRvKvIvtvSvCvKvOvTtEvlSvPvOvPvSvNvYvtvSvNvYv
 SvkRvKvIvAvtKvOvNvYvtvSvEvlSvGvRvIvtvLlADvDvSvFvHvEsvHvYvAvtOvAvtPvIvDvLlDvRvPv
 EsvMvKvHvRvSvVvEvlQvUvEvdvDvIvHvIvtvPvIvAvtDvGvHvIvQvOvFvIvAvtLlAvtRvKvIvCvDvGvDvEvlvAl
 RvSvYvIvtvPvKvOvEvlPvErvRvGvEvlTtCvSvGvCvIvMvRvEsvDvOvLttKvSvYvNvOvAvtPvIvNvAvtYvtvIvNvAvtGvS
 SvtvYvIvtvPvKvOvEvlDvIvNvDvRvNvDvLlAvtPvIvDvTt

ORIGIN

Query Match	88.0%;	Score 1633.6;	DB 9;	Length 1818;
Best Local Similarity	99.8%;	Pred. No. 4.9e-308;		
Matches 1636; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

QY	15	TCAAAATGATCACAAGTGTGGAAATACCCCTGGACGGGGCTTGGCCCTGTGTTTCCAA	74
Db	19	TCAAAATGATCACAAGTGTGGAAATACCTCTGGACGGGGCTTGGCCCTGTGTTTCCAA	78
QY	75	CCGCACATCCCGGAGTGCAGCTTACGTGTGGGGGAAAGCAGTGGCCATGGCAGACGGCA	134
Db	79	CCGCACATCCCGGAGTGCAGCTTTACGTGTGGGGGAAAGCAGTGGCCATGGCAGACGGCA	138
QY	135	GGGATACCCGGAACCCACGGAGGTGCAGATGAGCCAGCTGTGTCTGCCCTTGGCCACATCAA	194
Db	139	GGGATACCCGGAACCCACGGAGGTGCAGATGAGCCAGCTGTGTCTGCCCTTGGCCACATCAA	198
QY	195	CCAACGNGTGAGCTCAGCGCTCGGGCAGCGCTCAAGTGAATTGACACCAACGAGTTGCT	254
Db	199	CCAACGNGTGAGCTCAGCGCTCGGGCAGCGCTCAAGTGAATTGACACCAACGAGTTGCT	258
QY	255	GTCGCGGAGAGGCAACGCTGGCTGCCCTGTGTGCAGAGCTTCATGATGACATCTAATT	314
Db	259	GTCGCGGAGAGGCAACGCTGGCTGCCCTGTGTGCAGAGCTTCATGATGACATCTAATT	318
QY	315	TGAGCACACCATTAAGTGTGGACAAGTGTGAATATCAAGGCCAAGGTGAACGGGCGCTT	374
Db	319	TGAGCACACCATTAAGTGTGGACAAGTGTGAATATCAAGGCCAAGGTGAACGGGCGCTT	378
QY	375	CAACTCCAGATGAGAGTGGGACATCCAGGTGGCTGGAGAGCACTGTGCTGTGAAGAACA	434
Db	379	CAACTCCAGATGAGAGTGGGACATCCAGGTGGCTGGAGAGCACTGTGCTGTGAAGAACA	438

QY	435	GTGGAATGTGTGCAGAGCGTTGGCGACCTTGTCGGCCCGCCGAGAAGATCAACAAGTGA	494
Db	439	GTGGAAATGTGTGCAGAGCGCTTGGCACTTGTGTGGCCCGCCGAGAAGATCAACAAGTGA	498
QY	495	GCTTAAACAGATCAACGCCCGCGACAGAAAGAGAAAGATGAGACAGTGTGGCGCTGA	554
Db	499	GCTBAAGCAAGTCAACGCCCGCGACAGAAAGAGAAAGATGAGACAGTGTGGCGCTGA	558
QY	555	GCGCCGCGCATGCGCCTTGTCTATGCAAGACCATTAAGACCTCTGGCCAACTGCGC	614
Db	559	GCGCCGCGCATGCGCCTTGTCTATGCAAGACCATTAAGACCTCTGGCGCAACTGCGC	618
QY	615	CATTCAAGGCGCATCTGGAGAGCAACAGACTGATGCCCGCATGGTGGCCGCTGAGAAACCG	674
Db	619	CATTCAAGGCGCATCTGGAGAGCAACAGACTGATGCCCGCATGGTGGCCGCTGAGAAACCG	678
QY	675	TGTTGAGAGTGTGAGCTGCTCTGCGCTCCCAACGCAATATCAAGGAGCAACCTTTGG	734
Db	679	TGTGGAGAGTGTGAGCTGCTCTGCGCTCCCAACGCAATATCAAGGAGCAACCTTTGG	738
QY	735	GAGGCAGATCATGSCCTTGATGAGAAATGTGGCCACATTGACAGCCAGCCGCTTCGCG	794
Db	739	GAGGCAGATCATGSCCTTGATGAGAAATGTGGCCACATTGACAGCCAGCCGCTTCGCG	798
QY	795	TGCCCACTCTACGCTGAAAGGCCATTGAAATGTTCACCTTCCGAGGCCCTGCCAGTTCG	854
Db	799	TGCCCACTCTACGCTGAAAGGCCATTGAAATGTTCACCTTCCGAGGCCCTGCCAGTTCG	858
QY	855	CGACCGTCTGTGCTCAAAAGCCATCGTGAACAAATGCTTCAAAATAGCATGAGAGTGG	914
Db	859	CGACCGTCTGTGCTCAAAAGCCATCGTGAACAAATGCTTCAAAATAGCATGAGAGTGG	918
QY	915	CGTGTGCGTGAAGSCCTATCGCCAGAGAGCTGAGAACCCACCGGCGCACATCAACAGTGC	974
Db	919	CGTGTGCGTGAAGSCCTATCGCCAGAGAGCTGAGAACCCACCGGCGCACATCAACAGTGC	978
QY	975	CTTTATATGACCTTTTGTGCTCTGAGACAGATGACAGCCCAAGTTGCTGCTGATTCG	1034
Db	979	CTTTATATGACCTTTTGTGCTCTGAGACAGATGACAGCCCAAGTTGCTGCTGATTCG	1038
QY	1035	GCCCAAGCGCGGCGCATGTGTGAGCGCGCGGTACCGAGAGGCCAGTGCACAGAAAGAAATCG	1094
Db	1039	GCCCCAGCGCGGCGCATGTGTGAGCGCGCGGTACCGAGAGGCCAGTGCACAGAAAGAAATCG	1098
QY	1095	CCTGGAACAGAAATGATCATCGTGTCTCTGTAGACAGACAGAGGTGCCCTCTCCGTCCCTG	1154
Db	1099	CCTGGAACAGAAATGATCATCGTGTCTCTGTAGACAGACAGAGGTGCCCTCTCCGTCCCTG	1158
QY	1155	GAACTTATGCAACACAGGTGTACTGAGCTTAACAATATACGTCTCTCTTGAAGATGCTTGT	1214
Db	1159	GAACTTATGCAACACAGGTGTACTGAGCTTAACAATATACGTCTCTCTTGAAGATGCTTGT	1218
QY	1215	GAGCAAGAACATCTGGGTGTCTCTCGGACATCATAGGTCCGCTGTACATCTGGA	1274
Db	1219	GAGCAAGAACATCTGGGTGTCTCTCGGACATCATAGGTCCGCTGTACATCTGGA	1278
QY	1275	GGATGACAAGTTCCTCTCTCTTCCACATGAGATGGTGTGATGTGATGAGCCCAAGC	1334
Db	1279	GGATGACAAGTTCCTCTCTCTCTTCCACATGAGATGGTGTGATGTGATGAGCCCAAGC	1338
QY	1335	CTTCTGCTGCTCTCTGACCTTGCTCAGAGCCAGAGTGGGACAAAGCATACCGAGGCT	1394
Db	1339	CTTCTGCTGCTCTCTGACCTTGCTCAGAGCCAGAGTGGGACAAAGCATACCGAGGCT	1398
QY	1395	GGAAGTATGACAGCGTGAAGACAGAGAGAGCGCCATCTACACAGTCAACAGCCCTGCT	1454
Db	1399	GGAAGTATGACAGCGTGAAGACAGAGAGAGCGCCATCTACACAGTCAACAGCCCTGCT	1458
QY	1455	CGAAGTCAACAAGAGCCCAAGAGATTGCTGATCTCTGGCTCGAGAGCCGAAAGCTTGTGA	1514
Db	1459	CGAAGTCAACAAGAGCCCAAGAGATTGCTGATCTCTGGCTCGAGAGCCGAAAGCTTGTGA	1518
QY	1515	CAATGGAGCCCTATGTGATCGCGCTAGAGTGGGTCAAGCTTGCCCAACAACGAGAGAC	1574

Db	1519	CAATGGGAGCCCTTATCTATTCGGCGTGAAGGTGGGTCAAGCTGCGCCACACACGAGAGAC	1578
Qy	1575	GCCAGAGTACAGACGCGGAGAGACCTCTGTGCTCAGAGCTTCTGCTCTGGGCGCAGGGGGA	1634
Db	1579	GCCAGAGTACAGACGCGGAGAGACCTCTGTGCTCAGAGCTTCTGCTCTGGGCGCAGGGGGA	1638
Qy	1635	CCAGCTGACCAAGTCGCT	1654
Db	1639	CCAGCTGACCAAGTATCT	1658
RESULT 4			
AX883384	AX883384	3082 bp	DNA
LOCUS	Sequence 18289 from Patent EP1074617.		linear
DEFINITION	AX883384		
ACCESSION	AX883384		
VERSION	AX883384.1	GI:40038285	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1	Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,...	
AUTHORS	1	Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,...	
TITLE	Primer for synthesizing full-length cDNA and their use		
JOURNAL	Research Association for Biotechnology (JP)		
FEATURES	1..3082		
SOURCE	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	83..1867		
	/note="unnamed protein product"		
CDS	/codon_start=1		
	/protein_id="CA891952.1"		
	/db_xref="GI:40038286"		
	/translation="MIQVGNLHRLGLASVPSNRTSRKSLALRAGNDSAMADGEGYRNF		
	TEVMSQVLPECHTNGELSVQLKIMIDTACLSERHAGCEVTASMDIYFEHT		
	ISVGQVNIKAQVRAFNSSMEVGIQVASEDLGSEKQNVKALATFARBEI RFVKI		
	KQIPIREBEKMSVVAERRMFLVADIT KOLLANALGIDESHRCMVAPEAKI		
	KVSEVLELPRPHAHQGTNRGGQIMAMKENVATTAASRLCAHPTTLAKIEFPHRGP		
	QVGRVLVKAIVNNAFHSMEVGVCEVAERAEATHRHINSAFMTFVLLADDPQ		
	LPMPQPGGERRERYEASARKIRLRKYIVLSQKTEVPLSVMDPSNQLYSNN		
	ESLHSLVAKNWVLSSEISQVRLYLIEDKFLSPHEMENVVDAQAQFLISLROR		
	EMDGLSVSEVLEQVDEVDATVHTSPALGGTROPDVLIASRRKPCDNDSPYIAL		
	RSYVLPHREPRERRGPTLCSGFLWRSESGQLKVSYYNATGEVLYVTNVAGLS		
	SEFYTFRACQFLDNRNDLAPSLQTL"		
ORIGIN			
Query Match	87.7%; Score 1628; DB 6; Length 3082;		
Best local similarity	99.7%; Pred. No. 5.7e-307;		
Matches 1631; Conservative	0; Mismatches 5; Indels 0; Gaps 0		
Qy	19	ATGATCCGAATGTGGAAATACCTTGCGACGGGGCTTGCCCTGTGTCTCCAAACGC	78
Db	83	ATGATCCGAATGTGGAAATACCTTGCGACGGGGCTTGCCCTGTGTCTCCAAACGC	142
Qy	79	ACATCCGGGAAGTCAAGCTTACGTGGGGGAAACGACATGCGATGGCAGACGGGCAAGGA	138
Db	143	ACATCCGGGAAGTCAAGCTTACGTGGGGGAAACGACATGCGATGGCAGACGGGCAAGGA	202
Qy	139	TACCGGAACCCCAAGAGGTGACATGAGCAGCGTGGTGTGCTGCTGCACCAACCA	198
Db	203	TACCGGAACCCCAAGAGGTGACATGAGCAGCGTGGTGTGCTGCTGCACCAACCA	262
Qy	199	CGTGTGAGCTGAGGCGTGGGACGCTGCTCAAGTGATGACACCAAGGCTTGCCTGTC	258
Db	263	CGTGTGAGCTGAGGCGTGGGACGCTGCTCAAGTGATGACACCAAGGCTTGCCTGTC	322
Qy	259	GCGAGAGACACGCTGGTGGCCCTGTGTCAACGTTCCATGAGATGACATATTTTGA	318

Db	323	GGGAGAGGACGCTGGGCTGCCCTGTGTGACAGCTCCATGGATACATCTATTTTGAG	382
Qy	319	CACACCATTAAGTGTGGACAAGTGGTAATATCAAGGCCAAGGTGAACCGGGCTTCAAC	378
Db	383	CACACCATTAAGTGTGGACAAGTGGTAATATCAAGGCCAAGGTGAACCGGGCTTCAAC	442
Qy	379	TCCAGCATGAGTGGGACATCCAGGTGGCTGGAGGACCTGTGCTCTGAGAGAGAGTGG	438
Db	443	TCCAGCATGAGTGGGACATCCAGGTGGCTGGAGGACCTGTGCTCTGAGAGAGAGTGG	502
Qy	439	AATGTGTGCAAGGCTTGGCCA CTTTCTGTGGCCCGCCGAGAGATCAACAAAGTGAAGT	498
Db	503	AATGTGTGCAAGGCTTGGCCA CTTTCTGTGGCCCGCCGAGAGATCAACAAAGTGAAGT	562
Qy	499	AAGCAGATCAAGCCCGGACAGAAAGAGAGAAATGAGACAAGTGTGGCGCTGAGCGC	558
Db	553	AAGCAGATCAAGCCCGGACAGAAAGAGAGAAATGAGACAAGTGTGGCGCTGAGCGC	622
Qy	559	CGGCGCATGCGCTGTCTATGACAGACATCAAGGACCTCTGGGCCAACTGGCCCAAT	618
Db	623	CGGCGCATGCGCTGTCTATGACAGACATCAAGGACCTCTGGGCCAACTGGCCCAAT	682
Qy	619	CAGGCGCATCTGGAGAGACAGACATCTAGCCGATGTGCTGAGAGAGACCGGTGTG	678
Db	683	CAGGCGCATCTGGAGAGACAGACATCTAGCCGATGTGCTGAGAGAGACCGGTGTG	742
Qy	679	GAGAGTGTGAGAGCTGTGCTGCTCCCAAGCCCAATCAACAGAGGCAACCTTTGGGGCC	738
Db	743	GAGAGTGTGAGAGCTGTGCTGCTCCCAAGCCCAATCAACAGAGGCAACCTTTGGGGCC	802
Qy	739	CAGATCATGAGCTGTGATGAGAGATGTGGCAACATTTGAGGACGAGCGGCTGCGGTGC	798
Db	803	CAGATCATGAGCTGTGATGAGAGATGTGGCAACATTTGAGGACGAGCGGCTGCGGTGC	862
Qy	799	CACCTTACGCTGAGAGGCAATGTAATGTTCACTTCCAGAGCCGCTCCCAAGTCCGCGAC	858
Db	863	CACCTTACGCTGAGAGGCAATGTAATGTTCACTTCCAGAGCCGCTCCCAAGTCCGCGAC	922
Qy	859	CGTGTGTGTCTCAAGGCAATGTAATGTTCACTTCCAGAGCCGCTCCCAAGTCCGCGAC	918
Db	923	CGTGTGTGTCTCAAGGCAATGTAATGTTCACTTCCAGAGCCGCTCCCAAGTCCGCGAC	982
Qy	919	TGCGTGAAGGCTTATCGCCAGAGGCTGAGACCCAGCGGCGCAATCAACAGTGCCTT	978
Db	983	TGCGTGAAGGCTTATCGCCAGAGGCTGAGACCCAGCGGCGCAATCAACAGTGCCTT	1042
Qy	979	ATGACCTTTGTGTCTGAGACGAGATGACAGCCCGAGTTGCTGCTGAGTTGCGGCC	1038
Db	1043	ATGACCTTTGTGTCTGAGACGAGATGACAGCCCGAGTTGCTGCTGAGTTGCGGCC	1102
Qy	1039	CAGCCCGGCGATGTGGAGCGCGCGATGACCGAGAGGCTATGCCAAGAAAGATCGGCTG	1098
Db	1103	CAGCCCGGCGATGTGGAGCGCGCGATGACCGAGAGGCTATGCCAAGAAAGATCGGCTG	1162
Qy	1099	GACAGGAAGTACATGCTGTCTGTAAAGACAGAGAGTGGCCCTGCTGCTGCGGAGC	1158
Db	1163	GACAGGAAGTACATGCTGTCTGTAAAGACAGAGAGTGGCCCTGCTGCTGCGGAGC	1222
Qy	1159	CTTACGCAACGAGTGTACCTGAGCTAACATACGCTCTCTTGAAGATCTTGTGGCC	1218
Db	1223	CTTACGCAACGAGTGTACCTGAGCTAACATACGCTCTCTTGAAGATCTTGTGGCC	1282
Qy	1219	AAGGACAACTGGGAGTGTCTCTCGAGATCACTAGAGTGGCCCTGCTGCTGAGAGAT	1278
Db	1283	AAGGACAACTGGGAGTGTCTCTCGAGATCACTAGAGTGGCCCTGCTGCTGAGAGAT	1342
Qy	1279	GACAAAGTTCCTCTCTCTCAATGAGATGAGTGGTGCATGTGAGTGCAGCCAGGCTTC	1338
Db	1343	GACAAAGTTCCTCTCTCTCAATGAGATGAGTGGTGCATGTGAGTGCAGCCAGGCTTC	1402
Qy	1339	CTGTGTCTTGCAGCTGTGCTGAGAGGCTGAGAGTGGGCAAGCACTACCGAGGCTGAG	1398
Db	1403	CTGTGTCTTGCAGCTGTGCTGAGAGGCTGAGAGTGGGCAAGCACTACCGAGGCTGAG	1462
Qy	1399	CTAGTGCAGCAGTATGAG	1458
Db	1463	CTAGTGCAGCAGTATGAG	1522
Qy	1459	GGTCAACAAAGCCCGAGAGACTTGTGATCTTGTGATCTTGTGATCTTGTGATCTTGT	1518
Db	1523	GGTCAACAAAGCCCGAGAGACTTGTGATCTTGTGATCTTGTGATCTTGTGATCTTGT	1582
Qy	1519	GGGAGCCCTATGATCATGCGCTGAGAGTGGTCAAGCTGCGCCACACAGAGAGAGAG	1578
Db	1583	GGGAGCCCTATGATCATGCGCTGAGAGTGGTCAAGCTGCGCCACACAGAGAGAGAG	1642
Qy	1579	GAGTACAGACGCGGAG	1638
Db	1643	GAGTACAGACGCGGAG	1702
Qy	1639	CTGACCAAGTGTGCT 1654	
Db	1703	CTGACCAAGTGTGCT 1718	
RESULT 5			
BD160291			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
1 (bases 1 to 3082)			
REFERENCE			
Ota, T., Iwagaki, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,			
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.			
Primer for synthesizing full-length cDNA and use thereof			
Patent: JP 2002191363-A 15134 09-JUN-2002;			
JOURNAL			
HEPILX RESEARCH INSTITUTE			
OS Homo sapiens (human)			
PN JP 2002191363-A/15134			
PD 09-JUN-2002			
PF 28-JUN-2000 JP 2000280990			
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU			
PI SAITO,			
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,			
PI KEIICHI NAGAI, TETSUJI OTSUKI			
PC			
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/			
10,			
PC C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC			
Primer for synthesizing full-length cDNA and use thereof FH Key			
Location/Qualifiers			
FT CDS			
Location/Qualifiers			
1..3082			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
ORIGIN			
Query Match			
Best Local Similarity 99.7%; Pred. No. 5.7e-307;			
Matches 1631; Conservative 0; Mismatches 5; Indels 0; Gaps 0;			
Qy	19	ATGATCCAGAAATGCGGAATCACTGGAGAGGAGCTTGGCTGTCTGTCACACCGC	78
Db	83	ATGATCCAGAAATGCGGAATCACTGGAGAGGAGCTTGGCTGTCTGTCACACCGC	142
Qy	79	ACATCCGGAAGTACAGCTTACGTGCGGGAACAGACAGTGCATATGACAGGCGAGGGA	138
Db	143	ACATCCGGAAGTACAGCTTACGTGCGGGAACAGACAGTGCATATGACAGGCGAGGGA	202

CDS

```

/clone_1b="THY01"
/note="Cloning vector: pME18SFL3"
63..1867
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14734.1"
/db_xref="gi:10436032"
/translation="MIQNVNHLRGLASVFSNRTSRKSLRAGNDSAMADSEYRNP
TEVMSOLVLPCHNRGELISVGLLKKIIDDITACLSARHAGCPVATSMDDIYRHT
ISVGVNVIKAKVNRARNSMEVGIQVASEDLCEKQNMVCKALTFVAREITKVL
KQITPRIEEKMEHSVAERRRRLVADTIKDLANCAIQDLERDSCRNVPEKKT
QVESVELLPVHNAQNTFGQIMAMENVAITIASPLCAHPTLKA1EMEHFEGPS
QVDRVLKAIYVNAFKHSMVEGVCEVAERQAEHRHINSAFWT EYVLDADPOL
LPMIPQDGERRRYREASARKILDKRYIVSCQTEVPSVMPDNOYVLSNNY
SLKMLVAKNMVLSSEISQVRLYTLDEKELSPHMEVYVVDAAQALILSDLPOR
EMDKHRSVELVQVDEDDAIYHTYSPALGHTKPODVIILASRKPCKDNEDPYIAL
RSVTLPIRETPETHRGELSCFCLMEEGQDLTKVSYNNAATPEVLNVYTTNAGLS
SEFYTFRACEQFLIDNRNDLAPSLQTL"

```

ORIGIN

Query Match 87.7% Score 1628; DB 9; Length 3082;
 Best Local Similarity 99.7% Pred. No. 5,7e-307;
 Matches 1631; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

19 ATGATCCAGAAATGTCGAAATCACTGCGACGCGGCTTG3CCTGTGTCTCAACCGC 78
83 ATGATCCAGAAATGTCGAAATCACTGCGACGCGGCTTG3CCTGTGTCTCAACCGC 142
79 ACATCCCGGAAAGTCAGCTTACGTGCGGAGAAAGACAGTGCCTAGGCGAGCGAGG 138
143 ACATCCCGGAAAGTCAGCTTACGTGCGGAGAAAGACAGTGCCTAGGCGAGCGAGG 202
139 TACCGGAAAGCGGAGGATGAGTGAAGCAGTGTGCTGCTGCGACCAACCAACCA 198
203 TACCGGAAAGCGGAGGATGAGTGAAGCAGTGTGCTGCTGCGACCAACCAACCA 262
199 CGTGTGAGCTGAGAGCTGCGGAGCTGCTCAAGTGAATGACACCGAGCTTGCTGTC 258
263 CGTGTGAGCTGAGAGCTGCGGAGCTGCTCAAGTGAATGACACCGAGCTTGCTGTC 322
259 GCGGAGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318
323 GCGGAGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 382
319 CACACCATTAAGTGTGACAGTGTGATTAATCAAGGCAAGGTGAACCGGCGCTTGAC 378
383 CACACCATTAAGTGTGACAGTGTGATTAATCAAGGCAAGGTGAACCGGCGCTTGAC 442
379 TCCAGCATGAGAGTGGGATCCAGTGGCTCGAGAGACCTGTGCTGAGAGAGAGTGG 438
443 TCCAGCATGAGAGTGGGATCCAGTGGCTCGAGAGACCTGTGCTGAGAGAGAGTGG 502
439 AATGTGTGCAAGGCTTGCGCACTTGTGGCCCGCGAGAGATTAACCAAGGTGAAGTGG 498
503 AATGTGTGCAAGGCTTGCGCACTTGTGGCCCGCGAGAGATTAACCAAGGTGAAGTGG 562
499 AAGCAGATCAAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 558
563 AAGCAGATCAAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 622
559 CGGCGCATGCGCTTGTCTATGACAGACCAATCAAGGACCTCTGGCCCAATGCGCCAT 618
623 CGGCGCATGCGCTTGTCTATGACAGACCAATCAAGGACCTCTGGCCCAATGCGCCAT 682
619 CAGGCGCATGCGCTTGTCTATGACAGACCAATCAAGGACCTCTGGCCCAATGCGCCAT 678
683 CAGGCGCATGCGCTTGTCTATGACAGACCAATCAAGGACCTCTGGCCCAATGCGCCAT 742
679 GAGAGTGTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738
743 GAGAGTGTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 802
739 CAGATCATGCGCTTGTCTATGAGAGATGTGGCAACCAATGAGCGGAGCGGCTTGCGCTG 798

```

```

DB 803 CAGATCATGCGCTTGTCTATGAGAGATGTGGCAACCAATGAGCGGAGCGGCTTGCGCTG 862
QY 799 CACCTTACGCTTGAAGGCAATGAAATGTTCCATTCCAGAGCGCGCTGCCAGTCCGAGC 858
DB 863 CACCTTACGCTTGAAGGCAATGAAATGTTCCATTCCAGAGCGCGCTGCCAGTCCGAGC 922
QY 859 CGTCTGTGTCTTAAAGGCAATGTTGAACCAATGCTTCAACCAATGAGAGTGGGCTG 918
DB 923 CGTCTGTGTCTTAAAGGCAATGTTGAACCAATGCTTCAACCAATGAGAGTGGGCTG 982
QY 919 TCGGAGAGGCGCTTATGCGCAGAGAGGCTGAGACCCACCGGCGCACATCAAGTGCTT 978
DB 983 TCGGAGAGGCGCTTATGCGCAGAGAGGCTGAGACCCACCGGCGCACATCAAGTGCTT 1042
QY 979 ATGACCTTGTGTGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1038
DB 1043 ATGACCTTGTGTGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1102
QY 1039 CAGCCCGGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1098
DB 1103 CAGCCCGGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1162
QY 1099 GACAGGAAGTACATGCTGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1158
DB 1163 GACAGGAAGTACATGCTGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1222
QY 1159 CCTAGCAACAGGTGTACCTGAGCTCAATAAAGTCTCTCTCTGTAAGATGCTTGTGAGC 1218
DB 1223 CCTAGCAACAGGTGTGTACCTGAGCTCAATAAAGTCTCTCTCTGTAAGATGCTTGTGAGC 1282
QY 1219 AAGGCAACTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1278
DB 1283 AAGGCAACTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1342
QY 1279 GACAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1338
DB 1343 GACAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1402
QY 1339 CTGCTGTCTTGTGAGAGCTGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1398
DB 1403 CTGCTGTCTTGTGAGAGCTGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1462
QY 1399 CTAGTGACAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1458
DB 1463 CTAGTGACAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1522
QY 1459 GGTACACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
DB 1523 GGTACACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1582
QY 1519 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1578
DB 1583 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1642
QY 1579 GAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
DB 1643 GAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1702
QY 1639 CTGACCAAGTGTGCT 1654
DB 1703 CTGACCAAGTGTGCT 1718

```

RESULT 7
 AK091112
 LOCUS
 DEFINITION Homo sapiens cDNA FLJ33793 fis, clone CD34C2000443, weakly similar
 to Aey1-coA hydrolase.
 ACCESSION AK091112
 VERSION AK091112.1 GI:21749405
 KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
Ninomiya, K., Magatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,
Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, B., Omura, Y., Abe, K.,
Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,
Kikuchi, H., Yamashita, H., Mateno, K., Nakamura, Y., Sekine, M.,
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
Nagahara, K., Masuno, Y., Nagai, K., and Isogai, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2875)

TITLE Isogai, T. and Yamamoto, J.
AUTHORS Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1. 2875
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CD34C2000443"
/cell_type="CD34+ Cells"
/clone_lib="CD34C2"
/note="cloning vector: pME18SFL3-primary culture, CD34+
Cells"

ORIGIN
Query Match 78.1%; Score 1450; DB 9; Length 2875;
Best Local Similarity 95.0%; Pred. No. 2,7e-272;
Matches 1536; Conservative 0; Mismatches 10; Indels 70; Gaps 1;

39 TCACCTCGGACGCGGCTTGGCTCTGTCTCCACCGGACATCCGGAAGTACGCTT 98
169 TCCCATCTTTGGGGCTTGGCTCTGTCTCCACCGGACATCCGGAAGTACGCTT 228
99 ACCGTGCGGGAAACACAGTGCATGCGAGCGAGGAGATCCGGAACCCCAAGAGT 158
229 ACGTGGCGGGAACACAGTGCATGCGAGCGGAGGAGATCCGGAACCCCAAGAGT 288
153 GCAGATAGGCGAGCTGTGTCTGCTTCCACACCAACCAAGTGTGAGTGTGCG 218
289 GCAGATAGGCGAGCTGTGTCTGCTTCCACACCAACCAAGTGTGAGTGTGCG 348
219 GCAGCTCTCAAGTGAATGACACACGCGTGTGCTGCGCGGAGAGGACGCTGCG 278
349 GCAGCTCTCAAGTGAATGACACACGCGTGTGCTGCGCGGAGAGGACGCTGCG 388
279 CCCCTGTGTACACACTTCATGATGACATCTATTTGAGGACACCATTAAGTGTGCA 338
389 -----GTGTTGACCA 398
339 AGTGTGAATATCAAGGCCAAAGTGAACCGGGCTTCAATCCAGCATGAGGTGGCAT 398
399 AGTGTGAATATCAAGGCCAAAGTGAACCGGGCTTCAATCCAGCATGAGGTGGCAT 458
399 CCAGTGGCTTCGAGAGACCTGTCTTGAAGAGAGGAAATGTGCGAGGCTTGGC 458
459 CCAAGTGGCTTCGAGAGACCTGTCTTGAAGAGAGGAAATGTGCGAGGCTTGGC 518

459 CACCTTCGTGGCCCCCGGAGATCAACCAAGTGAAGCTGAAGACATCAACGCCGGGAC 518
519 CACCTTCGTGGCCCCCGGAGATCAACCAAGTGAAGCTGAAGACATCAACGCCGGGAC 578
519 AGAAGGAGGAAGATGAGACAGTGTGGCGGTGAAGCGCGCGCATGGCGCTTGTCTA 578
579 AGAAGGAGGAAGATGAGACAGTGTGGCGGTGAAGCGCGCGCATGGCGCTTGTCTA 638
579 TGCAGACACCATCAAGGACCTCTTGGCCCAATCGCATTCAGGCGCATGTGAGAGCAG 638
639 TGCAGACACCATCAAGGACCTCTTGGCCCAATCGCATTCAGGCGCATGTGAGAGCAG 698
639 AGACTGTAGCGGATGAGTGGCTGGCTGAGAGACCCGTGTGAGAGTGTGAGAGCTGCT 758
699 GCTTCCCGGACCAATCAACAGGGCAACACTTTGGGGCGCAGATCATGGCTGGATGGA 758
759 GCTTCCCGGACCAATCAACAGGGCAACACTTTGGGGCGCAGATCATGGCTGGATGGA 818
759 GAATGTGGCCACCATTCGACAGCGGCTCTGCGCTGCGCCACCTTAAGCTGAAGGCAT 818
819 GAATGTGGCCACCATTCGACAGCGGCTCTGCGCTGCGCCACCTTAAGCTGAAGGCAT 878
819 TGAATGTTCACCTTCGAGGCGGCTTCCAGGTGTGCGGACCGTGTGCTCAAGGCAT 878
879 TGAATGTTCACCTTCGAGGCGGCTTCCAGGTGTGCGGACCGTGTGCTCAAGGCAT 938
879 CGTGAACAAATGCTTCAACATATGACATGAGGAGTGTGGGTGAGGCGCTTACGCA 938
939 CGTGAACAAATGCTTCAACATATGACATGAGGAGTGTGGGTGAGGCGCTTACGCA 998
939 GAGAGCTGAGACCCACCGGCGGACATCAACAGTGCCTTTATGACCTTTGTGATCTGGA 998
999 GAGAGCTGAGACCCACCGGCGGACATCAACAGTGCCTTTATGACCTTTGTGATCTGGA 1058
999 CGAGATGACCAAGCCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1058
1059 CGAGATGACCAAGCCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1118
1059 GCGGTACCGAGGCGGACGAGTGCAGAAAGAGTCCGCTGAGAGGAGTACATGCTGTC 1118
1119 GCGGTACCGAGGCGGACGAGTGCAGAAAGAGTCCGCTGAGAGGAGTACATGCTGTC 1178
1119 CTGTAAACAGACAGAGGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1238
1179 GAGCTAACATPAAGTCTCTCTCTTGAAGATGCTTGTGGCAAGCAACTGGTGTCTC 1238
1239 GAGCTAACATPAAGTCTCTCTCTTGAAGATGCTTGTGGCAAGCAACTGGTGTCTC 1298
1239 CTGGAGATCAATGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1298
1299 CTGGAGATCAATGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1358
1359 CATGAGATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1418
1359 TCAGAGGCGAGAGTGGAGCAAGCACTACCGAGAGTGTGATGATGATGATGATGATG 1478
1419 TCAGAGGCGAGAGTGGAGCAAGCACTACCGAGAGTGTGATGATGATGATGATGATG 1478
1419 GAGAGGCGCATTAACAGAGTCAACAGGCGCTGCTGAGAGTCAACAAAGCCCAAGA 1478
1479 GAGAGGCGCATTAACAGAGTCAACAGGCGCTGCTGAGAGTCAACAAAGCCCAAGA 1538
1479 CTTGTGATCTGTGCTGAGGCGGAGCTTGTGACATGAGGAGCCCTTATGTCTATGCT 1538
1539 CTTGTGATCTGTGCTGAGGCGGAGCTTGTGACATGAGGAGCCCTTATGTCTATGCT 1598
1539 GCTGAGGTCTGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1598

Db	1599	GCTGAGCTCGGTCACGCTGCTCCACACACCGAGAGCGCCAGATGACAGCGGAGAGC	1658
OY	1599	CCTGCTCAGGCTTCTGCTCTGCGCGGAGGGGACCAAGCTGACCAAGTGTCT	1654
Db	1659	CCTGCTCAGGCTTCTGCTCTGCGCGGAGGGGACCAAGCTGACCAAGTGTCT	1714
RESULT 8			
AF416923			
LOCUS	AF416923	2678 bp	mRNA linear ROD 13-DEC-2001
DEFINITION	Mus musculus brown fat inducible thioesterase 2 mRNA, complete cds.		
ACCESSION	AF416923		
VERSION	AF416923.1	GI:17646240	
KEYWORDS			
SOURCE			
ORGANISM	Mus musculus (house mouse)		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 2678)		
JOURNAL	Adams, S.H., Chui, C., Schilbach, S.L., Yu, X.X., Goddard, A.D.,		
REFERENCE	Grimaldi, C., Lee, J., Dowd, P., Colman, S. and Lewin, D.A.		
AUTHORS	BRIT, a Unique Acyl-CoA Thioesterase Induced in Thermogenic Brown		
TITLE	Adipose Tissue. Cloning, Organization of the Human Gene, &		
JOURNAL	Assessment of a Potential Link to Obesity		
REFERENCE	Biochem. J. (2002) In press		
AUTHORS	2 (bases 1 to 2678)		
TITLE	Adams, S.H., Chui, C., Schilbach, S.L., Yu, X.X., Goddard, A.D.,		
JOURNAL	Grimaldi, C., Lee, J., Dowd, P., Colman, S. and Lewin, D.A.		
REFERENCE	Submitted (07-SEP-2001) Endocrinology, Genentech, Inc., 1 DNA Way,		
AUTHORS	South San Francisco, CA 94080, USA		
FEATURES			
SOURCE	Location/Qualifiers		
	1..2678		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="FVB/N"		
	/db_xref="taxon:10090"		
	/chromosome="4"		
	54..1838		
	/note="BFIT2; acyl-CoA thioesterase; acyl-CoA hydrolase;		
	similar to Homo sapiens BFIT2"		
	/codon_start=1		
	/product="brown fat inducible thioesterase 2"		
	/protein_id="AAL0093.1"		
	/db_xref="GI:17646241"		
	/translation="MIONGVNHLRSGFASMFNSRKSISHPESGDPMTAGSEGR		
	NPTAEVQSVLTPCHTHRGELSLGILGXMDTPACI.SARHAGCPVTSMDIYD		
	HTLSVGVNNTKAVNAPFNSMVGQVSVSEDCSKNSVCALATFYAHBELSKY		
	KLKVAVLPEBEKTEHGVAAERRRRLVADTIDKLTHCVIIDDLDKDSNNVPARK		
	TRVESVLEVPFANHQANTFGQIMAMENVAIIASRI,CHAPITKA,EMHFIRP		
	SOVEDRLVKAIVNNAFKHMEVGVCEAARQEAETRRHINSFMTFVLDODQO		
	KLPIRROPGEGERREYEAARKIRIDRYLVSCKAEVALSVPMDSOVLYSYN		
	VSLIKTMKDQNWYLSEIVLSVRLVITLEDPLSHLEMYNVNDAAVQFQLSLRRP		
	EMDHYSVLELVQGVDEDDAIYHISFALSNGRKPQDFVLASRRFCNDGDPVIL		
	RSVTLPHHETPEYOREETLCSGFLMRBEGDMTKYSYNQATPPFLNVTYVSGHS		
	SEFYNTKACESEFLDNRLDAPSLQTL"		
ORIGIN			
	66.2%; Score 1230; DB 10; Length 2678;		
	Best local Similarity 85.8%; Pred. No. 1.9e-229;		
	Matches 1404; Conservative 0; Mismatches 220; Indels 12; Gaps 3;		
OY	18	AATGATCCAGATTCGGAAATCACTGCGACGCGGCTTGCTGTGTTCCCAACG	77
Db	53	AATGATTCAGATGTGGGCAACCACTTGCGAAGGGCTTGCTGTGTTCTTAATCG	112
OY	78	CACATCCCGGAGTCAAGCTTACGTGC-----GGAAAGCAAGTCCCATGGCAGCG	131
Db	113	CACATCCCGAAGTCATCTCCATCCGAGTCTGGAGACCCCTCTACATAGGAGAGG	172
OY	132	CGAGGATACCGGAACCCACGAGGTGCGAGATGACCACTGTGTCTGCCCTGCCACAC	191

Db	173	TGAGAGATACCGGAAACCCACCGAAGTGCAGATGAGCAACGCTGGTACTGCCCTGCAAC	232
QY	192	CAACCAAGCTGTGAAGCTGAGCGTGGGAGCGTGTCTCAAGTGGATTTGACCAACAGGCTTG	251
Db	223	CAACCAAGCTGGGGAGCTGAGGATTTGACAGTGTCTCAAGTGGATTCGACACCAACGCTTG	292
QY	252	CCTGTCCGGAGAGAGGACAGCGTGGCCCTGTGTCAACGCTTGCATGATGACATCTA	311
Db	293	CCTATCACCGGAGAGGCAATGCTGGCTGTCCCTGCGTCAACGCTCTATGATGACATCTA	352
QY	312	TTTTAGACACCCATTAGTGTGGAAGAATGTGTAATTCAGAGCCAGAGTGGAAATCCGGC	371
Db	353	CTTGACCAATTCATTAGTGTGCGCAAGTGGTGAATTCAGAGCCAGAGTGGAAATCCGGC	412
QY	372	CTTCAACTCCAGCATGAGATGGGCGATCCAGGTGCCTCGAGAGCCGTGTCTCGAGAA	431
Db	413	CTTCAACTCCAGCATGAGAGTGGGAATCAAGTGTCTCTGAGAGATCTGTGTCTTGAGAA	472
QY	432	GCAGTGAATGTGTGCAAGAGCCTTGGCCACTTGTGTGGCCCGCGAGAGATCAACCAAGT	491
Db	473	GCAGTGAAGTGTGTGCAAGAGCCTTGGCCACTTGTGTGGCCCGCGAGAGATCAACCAAGT	532
QY	492	GAACTGTAAGAGATCAACGCCCGGACAGAAAGAGAAAGATGAGAGCAAGTGTGGCGGC	551
Db	533	GAACTGTAAGAGATCAATCCATTGACCGAGAGAGAAAGACTGAACATGGAGTGGCGGC	592
QY	552	TGAGGCGCGGCGCATGCGCTTGTCTATGACAGCAACCATCAAGAGCCCTCTGGCCACTG	611
Db	593	TGAGGCGCGGCGTATGCACTGTGTCTATGACAGCAACCATCAAGATCTCTTAACCACTG	652
QY	612	CGCAATTCAGGCGCATCTGAGAGACAGACTGTAGCCGATGGTGGCGGCTGAGAAAC	671
Db	653	TGTATCCAGAGACATTTGGA---CAAGGACTGCAGCAATATGGTGGCCAGCGAGAAAGAC	709
QY	672	CCGTGTGAGAGTGTGAGCTGTGTCTTGGCTCCCAAGCCCATATCACAGGGCAACACTT	731
Db	710	CCGATGTGAGAGTGTGAGCTGTGTCTTGGCTCTCTCAAGCCCATATCAAGGCAATACCTT	769
QY	732	TGGGGGGCAGATCAATGAGCTGAGTGAAGAAATGTGGCCACCATTCGAGCCAGCCGACTG	791
Db	770	CGGGGACAGATCAATGAGCTGTGTGAGAGAAATGTGGCCACCATTCGAGCCAGCCGACTG	829
QY	792	CGGTGCCCACTTACGCTGAAAGGCAATTTGAATTTCCACTTCCGAGGCCGCTCCAGGT	851
Db	830	TCAGGCCCACTTACGCTGAAAGGCAATTTGAATTTCCACTTCCGAGGCCGCTCCAGGT	889
QY	852	CGGAGACCGTGTGTGTCTCAAGGCACTGTGAAACATATGCTTTCAACATAGCATGAGAGT	911
Db	890	GAGGAGACCGTGTGTGTCTCAAGGCACTGTGAAACATATGCTTTTAAAGCACACATGAGAGT	949
QY	912	GAGGAGTGCATGAGAGGCTATCGCCAGAGAGGCTGAGAGCCCAACGCGCGCCACATCAACAG	971
Db	950	GAGTGTGTGTGTGAGAGGCTATCGCCAGAGAGGCTGAGAGCCCAACGCGCGCCACATCAACAG	1009
QY	972	TGCTTTATGACCTTTGTGTCTTGAACGAGATGACAGCCCAAGTTGTCTGCCCTGAT	1031
Db	1010	CGCTTTATGACCTTGTGTCTTGAACAAAGATGACAGGCTCAGAAAGCTGCCCTGAT	1069
QY	1032	TGGGCCCCAGCCCGGCGATGTGAGCGGCGGTATCCGAGAGCCAGTGTCCAGAAAGAAAT	1091
Db	1070	TGCTGCCAGCTGGAAGGAGGTGAACGGCAATACGAGAGCCAGTGTCCAGAAAGAAAT	1129
QY	1092	CCGCTCTGAGAGAAATACATCGTGTCTTGAAGCAACAGAGGTGGCCCTCTCCGCTCC	1151
Db	1130	CCGCTCTGAGAGAAATACCTTGTGTCTCTTGAAGCAACAGAGGTGGCCCTCTCTGCTCC	1189
QY	1152	CTGGGACCTTAGCAACAGGCTGATCCTGAGCTACATTAAGTCTCTCTTGAAGATGT	1211
Db	1190	CTGGGACCTTAGCAACAGGATATCTGAGCTACTAACAGTGTCTCTGAGAGAGCT	1249
QY	1212	TGTGGCCAAAGACAACTGGGTGTGTCTCTGGAAGATCAATCAAGTCCGCTGTACACTCT	1271
Db	1250	CATGGCCAAAGACAACTGGGTGTGTCTCTGGAAGATCAAGGAGGTGTCCGCTGTACACTCT	1309

Qy		1272	GGAGGATGACAAAGTTCCTCCTCCCTCCCACTGGAGATGGTGATGCATGGACGCCA	1331
Db		1310	AGAGAAGGAC---TTCCCTCTCTTCACTTGAATGGTGTAAATGGATGCGGCCA	1366
Qy		1332	GAGCCTTCCTGCTGCTCGACACTGCTCGAGAGCCAGAGTGGGACCAAGCAACTACCGAG	1391
Db		1367	GGTCTTTTCAGTCTGCTGTCAGACTCTGCCAGSAGACAGAGTGGGACCAAGCATTCGCGAG	1426
Qy		1392	CGTAGGACTAGTGCAGACAGGTAGA CGAAGACGACGCCATCTAACAGTCACACGCTCTGC	1451
Db		1427	TGTGAGACTGGTGCAGCAAGTGAATGAGATGAGCGCATCTAACAGTCATCACGCCGCG	1486
Qy		1452	CCTGCGAGGTCAKCAAAGCCCCGAGACTTGTGATCTTGGCTCTCGAGCGGAGCCTTG	1511
Db		1487	CCTGAGGGGAGAACCAAGCCCCAGACTTGTGATCCTGGCTCTCGAGCGGAGCCTTG	1546
Qy		1512	TGACAAATGGGGAGACCCTATGTCATCGAGCTGAGTGGGTACGCGTGGCCACACACCGAGA	1571
Db		1547	TGACAAATGGGGAGACCCTATGTCATTTGCCCTTAGTGGTCAACGCTGGCCACGACCATGA	1606
Qy		1572	GACGCCAAGATACAGACGCGGAGAGACCCCTTGCTCAGGCTTTCCTCTTGCGCGAGGG	1631
Db		1607	GACACCGGAATATCAACAGTGGGAGAGCTCTCTGTTCAAGCTTCTGTGTGTGGCGTGAAGG	1666
Qy		1632	GGACCAAGCTGACCAAG 1647	
Db		1667	GGACCAAGTGAATAAG 1682	
RESULT_9		BC042492	5606 bp mRNA linear ROD 07-OCT-2003	
LOCUS		Mus musculus thioesterase, adipose associated, mRNA (cDNA clone		
DEFINITION		MGC:25974 IMAGE:4240966), complete cds.		
ACCESSION		BC042492		
VERSION		BC042492.1 GI:27503494		
KEYWORDS		MGC.		
SOURCE		Mus musculus (house mouse)		
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS		Strausberg,R.L., Feltingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusik,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Uebelin,T.B., Toshitsuki,S., Canciani,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abrahamson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Wooley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huliy,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickinson,M.C., Rodriguez,A.C., Gilwood,J., Schmutz,J., Myers,R.M., Butlerfield,V.S., Krzywinski,M.I., Skalska,U., Smalls,D.E., Scherter,A., Schein,J.E., Jones,S.J., and Marra,M.A.		
TITLE		Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE		22388257		
PubMed		12477932		
REFERENCE		2 (bases 1 to 5606)		
AUTHORS		Strausberg,R.		
TITLE		Direct Submission		
JOURNAL		Submitted (02-JUN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
USA				

REMARK
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louesged, H., Kowis, C.R., Speed, A.J., Martin, R.G., Muzny, D.M., Nanaev, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAC plate: 31 Row: f Column: 3
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22165403.

FEATURES
 source
 1..5606
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:25974 IMAGE:4240966"
 /tissue_type="Kidney, normal. 5 month old male mouse."
 /clone_lib="NCI CGAP_Kid14"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1..5606
 /gene="Thba"
 /note="synonyms: 1110020M10Rik, Bf1t, Bf1TL, MGC25974, Thba1"
 /db_xref="LocusID:329910"
 /db_xref="MGI:1913736"
 51..1835
 /codon_start=1
 /product="thioesterase, adipose associated"
 /protein_id="AAH42492.1"
 /db_xref="GI:27503495"
 /db_xref="LocusID:329910"
 /translation="MTQVNGNHARGRFASMFSTSRKSIHPESGDPMTABEGGYV
 NPTVQMSQIVLPCHTNHREGLISGILLKMTDTTACISAERHACPCPTASMDIYF
 HTISVGQVNIKAKVNRAFNFSMEVGQVASEDCSKQSWSCALATFVAHRELSK
 KLVQVPLTEEEKTEHGVAARERMRRLVYADITDILTLTCVITQDDLDKCCNMPAEK
 TRVSEVLTPEPHNHGANTREGGQIMAMERVATIASRLCHAPTLKAIMFHRGGRG
 SQGDRLVTKAIVNAPFKSMENVGVCEAVROEATRRRHNSAPMTFVLDKDDPD
 KLTPIRPQSGEGRYRERASARKIRLDRKTVLSCKAVALSVPMDSNOVYLSTYY
 VSSIKTLMDKNWVLSEISERVLYLLEEDFLSHLEWVAVDAQVFLQSLSDLRK
 EMKHVRSVLEVOVEDDDAIYHVISPALSGNTKPOFVILASRRKPCDNDGPDYIA
 RSVLTLPHEHTEPVOREGELTCSGFCILMRGDPQMTKVSYYNQTGFLNVYTTNVSGL
 SEFPNPKACSEFLDNRNDLAPSLQTL"

misc_feature
 189..5620
 /note="COG1607; Region: Acyl-CoA hydrolase [lipid
 metabolism]"
 misc_feature
 684..1142
 /note="COG1607; Region: Acyl-CoA hydrolase [lipid
 metabolism]"
 misc_feature
 1203..1709
 /note="START; Region: in STAR and phosphatidylcholine
 transfer protein"
 /db_xref="CDD:smart00234"

ORIGIN
 Query Match 66.2%; Score 1230; DB 10; Length 5606;
 Best Local Similarity 85.8%; Pred. No. 1.8e-229;
 Matches 1404; Conservative 0; Mismatches 220; Indels 12; Gaps 3

Db	50	AAATGATTGAGATGTGGGCAACCACTTGGAGAGGGGCTTCGCTCTAATGTTCTTAATCG	109
QY	78	CACATCCGGAAAGTCAGCCTTAAGTACG-----GGAAACGACAGTGCACATGGCAGACGG	131
Db	110	CACATCCCGGAAGTCATCTCCATCCGAGAGTCGAGAACCCCTCAACATGGCAGAGGG	169
QY	132	CGAGGGATATCCGGAAACCCCAAGGAGGTGCAGATGACCCAGCTGTGTCTGCTCCCAAC	191
Db	170	TGAAGATACCGGAACCCCAAGAGGTGCAGATGACCCAGCTGTGTCTGCTCCCAAC	229
QY	192	CAACCAACCTGTGTGAGCTGAGCGTCGGGAGCGCTCAAGTGGATGACACCAAGGCTTG	251
Db	230	CACCAACCTGTGGGAGCTGAGCATTTGGAAGTTGCTCAAGTGGATGACACCAAGCTTG	289
QY	252	CCTGTCCCGGAGAGGCAAGCTGTGCTGCCCTGTGTACAGCTTCCATGATGACATCTA	311
Db	290	CCTATCAGGAGAGGCAAGCTGTGCTGTCCCTGCGCTCAGCCTCTAATGATGACATCTA	349
QY	312	TTTTAGGCACACCTAATAGTTGTGACAAATGGTGAATTAACAGGCCAAGGTACCGGGC	371
Db	350	CTTCGACCAATACCATTAGTGTGTGGCCAAATGGTGAATTAACAGGCCAAGGTACCGGGC	409
QY	372	CTTCAACTCCACCATTTGAGAGGTGGGCATCCAGGTGGCCTCGAGAGACCTGTGCTCGAGAA	431
Db	410	CTTCAACTCCACCATTTGAGAGGTGGGAATCCAGGTGTGCTGTGAGGATCTGTGCTCTGAGAA	469
QY	432	GCAATGGAATGTGTGCAAGGCTTGGCCACTTGTGGCCGCGCAGAGATCAACCAAGT	491
Db	470	GCAATGGAATGTGTGCAAGGCTTGGCCACTTGTGGCCCACTTGTGGCCCAAGGT	529
QY	492	GAAAGTCGAAGCAGATCAAGCGCCGGACAGAAAGAGAAATGAGACACAGTGTGGCGGC	551
Db	530	GAAAGTCGAAGCAGGTATCTCCATTACCGAGAGAGAAAGACTGAACATGGGGTGGCGGC	589
QY	552	TGAGGCGCGCGCATGCGCTTGTCTATGACAGACACATCAAGGACCTTCCTGGCCACATG	611
Db	590	TGAGGCGCGCGCATGCGCATGTGTGACAGACACATCAAGGACCTTCCTGGCCACATG	649
QY	612	CGCCATTGAGGGCGATCTTGAGAGCAGAGACTGTAGCGCATGTGTGCGGCTGAGAGAC	671
Db	650	TGTCATTCAGGACGATTTGGA---CAAGGACTGACGCAATATGATGCCAGCGAGAAAGAC	706
QY	672	CCGTGTGAGAGATGTGTGAGCTGTCTGTCTCCACGCGCAATCAACAGGCAACACTT	731
Db	707	CCGATGTGAGAGTGTGTGAGCTGTGTCTCTCTCCTCACCCCAATCAATCAGGCGCAATCTT	766
QY	732	TGGGGGCGAGATCATGGGCTGGATGGAAGATGTGGCCACATGTGAGGCAAGCGGCTGTG	791
Db	767	CGGGGGAAGATCATGGCTGTGATGGAAGATGTGGCCACATGTGAGGCAAGCGGCTGTG	826
QY	792	CCGTGCCCACTTACGCTGAAGAGCCATTGAATGTTTCCATCTTCCAGGCGCGGTCCAGAT	851
Db	827	TCAAGCCACCTAAGGCTCAAGAGCCATGAGATGTTTCAATTTCCGAGGCCGTGTCAAGT	886
QY	852	CGGCGCACCTGTGTGTGCTCAAAAGCATGTGAACAATGCTTTCAAACTAGCATGAGAGT	911
Db	887	GGGGGACCGTGTGTGTGCTCAAGGCGCATGTGAATTAACCTTTTAAGCACAGCATGAGGT	946
QY	912	GGGCGTGTGCGTGAAGGCGCTATGCGCAGAGAGGTGAGCCACCGGGGCGCATCAACAG	971
Db	947	GGGTGTGTGTGAGAGGCTACCGCAGAGAGCTGAGACCGACGCGCGGCATCAACAG	1006
QY	972	TGCTTTATGACCTTGTGTGCTTGTGACGCAAGATGACAGGCCCATGTTGTGCTGCTGAT	1031
Db	1007	CGCTTTATGACCTTGTGTGCTTGTGACCAAGATGACAGGCCCATGTTGTGCTGCTGAT	1066
QY	1032	TCGGCCCCAGCTCCGGCAGATGTGTGACGCGCGGTACCGAGAGGCGCATGTGCAGAAAGAT	1091
Db	1067	TCGTGCCACAGCTCGAGAGGGGTGAACGCGCATCCAGAGAACCTGCTCAGGAAGAGAT	1126
QY	1092	CCGCTGGAGCAGGAAGTATCATCGTGTCTGTAAACAGACAGAGGTGCCCTCTCCGTCCC	1151

Db		1127	CCGCTGGACAGGAATAACTTGTGTCTCTTTAAAGCAGGCGAAGTGCCCTGTCTGTCCC	1186
OY		1152	CTGGAGACCCTTAGCAACAACGATGTAACCTAGACTAACATTAACGTCCTCTCTGAAAGATGCT	1211
Db		1167	CTGGGACCCCTTAGCAACAACGATTAACCTTAAGCTTAACAACTGTCTCTCTTGTMAAGACGCT	1248
OY		1212	TGTGCCAAGGACAACTGGGTGTCTGCTCTCGAGATCATGATCAAGTCCGCCTGTACACTCT	1271
Db		1247	CATGGCCAAAGACAACTGGGTGTCTGTCCGTGGAGATCAGCGAGGTCGCTGTACATCT	1306
OY		1272	GGAGGATGTACAAGTTCCTCTCTCTTCCCATATAGAATATGGTGGTGCATGTGATATGACGCCCA	1331
Db		1307	AGAAAGSAC--TTCTCTCTCTTCACTTGTGAATGTGTGTAAATGTGATATGTCGCCCA	1363
OY		1332	GGCCTTCTGTGTCTCTCTGACCTTGCTCAGAGGCGAAGTGGGACAAAGCACTACCCGAG	1391
Db		1364	GGTCTTTAGGTGTGTGTATGAGACTGTGCGAGAGACAGAGTGGGACAAAGCAATTACCGGAG	1423
OY		1392	CGTGAAGCTTAATGACAGAGGTAAACAGAGACAGCGCACTTAACAAGTCAACAGCCCTGC	1451
Db		1424	TGTGAAGCTGTGTGACAGCAAGTGTATGTAGAGATGAGCGCCATCTAACACAGCTCACAGCCCGC	1483
OY		1452	CCTGGAGGATCACAAAGCCCCCGAGACTTGATATCTGGCCCTCGAGCGGAAAGCTTGG	1511
Db		1484	CTTGACCGGAAACCAAGCCCCAGAGATTTGTATATCTGTGACTCTGTAGCGGAAAGCTTGG	1543
OY		1512	TGACATATGGGAGACCCTATGTATCTCGGCTGAGTTCGGTCAACGCTGCCACACACACGAGA	1571
Db		1544	TGACATATGGGAGACCCTATGTATCTCGGCTGAGTTCGGTCAACGCTGCCACACACACATGA	1603
OY		1572	GACCCAGAGTACAGACGCGGAGAGACCTCTGTCAAGCTTCTTGCTCTGGCGCGAGGG	1631
Db		1604	GACACCGGAATACCAAGTGGGAGAGACTCTGTTCAGAGCTTCTGTGTGGCGTGAAGGG	1663
OY		1632	GGACCAAGCTGACCAAG	1647
Db		1664	GGACCAAGTGAAG	1679
RESULT 10				
AKI22350				
LOCUS	AKI22350	5415 bp	mRNA	linear ROD 15-MAR-2003
DEFINITION	Mus musculus mRNA for mKIAA0707 protein.			
ACCESSION	AKI22350			
VERSION	AKI22350.1	GI:28972356		
KEYWORDS	FLI CDNA.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 Okazaki,N., Kikuno,R., Ohara,R., Inamono,S., Aizawa,H., Yusa,S., Nakajima,D., Nigase,T., Ohara,O. and Koga,H.			
TITLE	Prediction of the coding sequences of mouse homologues of KIAA gene: II. The complete nucleotide sequences of 400 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries DNA Res. 10, 35-48 (2003)			
JOURNAL	2 (bases 1 to 5415)			
REFERENCE	Okazaki,N., Kikuno,R., Nigase,T., Ohara,O. and Koga,H.			
AUTHORS	Direct Submission			
TITLE	Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 229-0818, Japan (E-mail:kamatari@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)			
JOURNAL	The CREATE program supported by Japan science and technology corporation; cDNA full insert sequencing; Kazusa DNA Research Institute; cDNA library construction; clone selection and 5'- & 3'-end one pass sequencing.			
FEATURES	location/Qualifiers			
SOURCE	1. 5415 /organism="Mus musculus" /mol_type="mRNA"			

ORIGIN

```

/db xref="taxon:10090"
/clone="mbg07785"
/tissue_type="brain"
/dev stage="adult"
/note="vector:modified pBC SK-+"
1..5415
/gene="mKIA0707"
<1..1687
/gene="mKIA0707"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=2
/evidence=not_experimental
/product="mKIA0707 protein"
/protein_id="BAC65632.1"
/db_xref="GI:28972357"
/translation="pPTMARGEVNPTFEVMSOLVLPCHTHRGELSTGOLLKMDIT
TACLSABRHAGCCPVTRSMDDIYFDHTISGVYVNIKAKINRAFNSSMEVGIYVSED
LCEKQMSVCKALATFVAHRELSKVKIYVPLTEBKETHGVAERRRRLYADIT
KDLITHCVIDDIDKDCSNMVPARKTEVESEVLVLPANHQNCTFGQJTMAMENVA
TIAASRLCHAHPTLKAIEMFHFSPQVGRVLKAIYNNAFKHSMGVCVAROE
AETORRHINSAFMTEFVVLKDODPOKLPJTRPOGEBEERYREASARKJRLRKLYL
SCQAEVALISVPMDSNOYVLSYVNSLITLMAKQWVLSVERISVRLYLIEDFLS
PHLEMYNVDAVOYFOLISDLRRPPEMDKHYRSVELVQYDEDDALTYHVSPLASNT
KPDQFVLASRRKPCDMDPYVALRSVTLPTHETPEYDGRGELSGFCLMBGDQM
TKSYVNGATPGPLNVTVTNVSGLSSEFNTFACBSFLDNRNDLAPSLQTL"

Query Match      63.7%  Score 1182.2;  DB 10;  Length 5415;
Best Local Similarity 86.7%;  Pred. No. 3.7e-220;
Matches 1327;  Conservative 0;  Mismatches 198;  Indels 6;  Gaps 2;

QY 117 TGCCATGCGACAGCGCGAGGGATACCGGAAACCCACGAGGTGAGATGAGCCAGCTGT 176
DB 7 TACCATGCGAGAGGTGAGATACCGGAACCCACGAGGTGAGATGAGCCAGCTGT 66
QY 177 GCTGCCCTGCCACCAACCAACGCTGTGAGCTGAGCGTGGGCGAGCTGCTCAAAGTAT 236
DB 67 ACTGCCCTGCCACCAACCAACGCTGTGAGCGTGGGCGAGCTGAGCAGTTGCAAGTTGAT 126
QY 237 TGACACACGCGCTGCTGTCGCCGGGAGAGGACGCTGCTGCCCTGTGTCAAGCTTC 296
DB 127 CGACACACGCGCTGCTGTCGCCGGGAGAGGACGCTGCTGCCCTGTGTGTCAAGCTTC 186
QY 297 CATGATGACATCTATTGTCAGACACCATTAAGTGTGACCAAGTGTGATATCAAGGC 356
DB 187 TATGATGACATCTATTGTCAGACACCATTAAGTGTGACCAAGTGTGATATCAAGGC 246
QY 357 CAAGTGAACCGGCGCTTCAACCTCAGCATGAGAGTGGGCATCAAGTGTGCTGAGGA 416
DB 247 CAAGTGAACCGGCGCTTCAACCTCAGCATGAGAGTGGGCATCAAGTGTGCTGAGGA 306
QY 417 CTTGTGCTTGAAGACATGAGATGTGTGCAAGGCTTGGCCACTTTCGTGGCCGCCG 476
DB 307 TCTGTGCTTGAAGACATGAGATGTGTGCAAGGCTTGGCCACTTTCGTGGCCGCCG 366
QY 477 AGAGATGACCAAGGTGAAGTGAAGACATCAAGCCGCGGACAGAAAGAGAGATGGA 536
DB 367 GGAAGCTTCCAAAGTGAAGTGAAGACATCAAGCCGCGGACAGAAAGAGAGATGGA 426
QY 537 GCACAGTGTGGCGGCTGAGCGCGCGCATGCGCTTGTCTATGCAACACCATCAAGA 596
DB 427 ACATGGGCTGGCGGCTGAGCGCGCGCATGCGCTTGTCTATGCAACACCATCAAGA 486
QY 597 CTTCTGTGGCACTGGCGCATTTCAAGGGGATCTGAGAGCAAGAACTGAGCCGATGT 656
DB 487 TCTCTTAACCACTGTGTCAATCAAGAGATTTGAA--CAAGACTGCAAGAAATATGT 543
QY 657 GCCGCTGAGAAAGCCGCTGTGAGAGTGTGAGACTGTCTGCTGCCACGCAATCA 716
DB 544 GCCAGCCGAGAAAGCCGAGTGTGAGAGTGTGAGACTGTGCTGCTGCCACGCAATCA 603
QY 717 CCAAGGCAACACTTTTGGGGGCGCAGATCATGCGCTGATGAGATGTGGCCACCATTCG 776
```

```

DB 604 TCAGGGCAATACCTTCGGGGGAGACAGATCATGCGCTTGATGAAATGTGGCCACATTCG 663
QY 777 AGCCAGCCGCGCTTCGCGGTGCCACCCCTACCGTGAAGAGCCATTGAATGTTCCACTTCG 836
DB 664 AGCCAGCCGCGCTTCGTCACGCCACCCCTACCGTGAAGAGCCATTGAATGTTCCACTTCG 723
QY 837 AGGCGCGTCCAGAGTGGCGACCGCTGTGTCTCAAGGCCATCGTGAACATGCTTCA 896
DB 724 AGGCGCGTCCAGAGTGGCGACCGCTGTGTCTCAAGGCCATCGTGAACATGCTTCA 783
QY 897 ACATAGATGAGAGTGGCGCTGTGTCTCAAGGCCATCGTGAACATGCTTCA 956
DB 784 GCACAGATGAGAGTGGCGCTGTGTCTCAAGGCCATCGTGAACATGCTTCA 843
QY 957 GCAGCAATGAGAGTGGCGCTGTGTCTCAAGGCCATCGTGAACATGCTTCA 1016
DB 844 CGGCGCATCAACAGCGCTTTCATGACCTTGTGTCTGTGACCAAGATGACAGCTCA 903
QY 1017 GTTGTGCGCTGTGATTTGGGCCCAAGCCGCGATGTGAGCGGCGGTACCGAGGCGAG 1076
DB 904 GAAGCTGCTGATTTGATTTGCTGCTGAGAGAGGTGAACGCGCATACCGAAGCGAG 963
QY 1077 TGCCAGAAAGAAATTCGCTGTGACAGAGATGATCATGTGTCTGTGAAGACAGAGGT 1136
DB 964 TGCCAGAAAGAAATTCGCTGTGACAGAGATGATCATGTGTCTGTGAAGACAGAGGT 1023
QY 1137 GCCCTCTCGCTCCCTGAGACCCCTAGACCAACAGAGTGTACCTGATCAATAACGCTTC 1196
DB 1024 GGCCCTGTGTGTCCCTCGGACCTTACCAACAGAGTGTACCTGATCAATAACGCTTC 1083
QY 1197 CTCCTTGAAGATGTTGTGACCAAGCAACATGAGTGTGTCTGTGTGAGATCAATCAAGT 1256
DB 1084 CTCCTTGAAGATGTTGTGACCAAGCAACATGAGTGTGTCTGTGTGAGATCAATCAAGT 1143
QY 1257 CCGCTGTGTACCTGTGAGAGATGACAAAGTTCCTCTTCCACTGAGATGTGTGTGCA 1316
DB 1144 CCGCTGTGTACCTGTGAGAGATGACAAAGTTCCTCTTCCACTGAGATGTGTGTGCA 1200
QY 1317 TGTGATGAGAGCCGAGGCTTTCCTGCTGTGTGTGAGACCTGTGAGAGCCGAGAGTGGGA 1376
DB 1201 TGTGATGAGAGCCGAGGCTTTCCTGCTGTGTGTGAGACCTGTGAGAGCCGAGAGTGGGA 1260
QY 1377 CAAGCACTTACCGAGAGTGTGAGAGTGTGACAGAGTGTGACAGAGTGTGACAGAGTGTG 1436
DB 1261 CAAGCACTTACCGAGAGTGTGAGAGTGTGACAGAGTGTGACAGAGTGTGACAGAGTGTG 1320
QY 1437 CGTACCAAGCCCTGCTGAGAGTGTGACAGAGTGTGACAGAGTGTGACAGAGTGTG 1496
DB 1321 CGTACCAAGCCCTGCTGAGAGTGTGACAGAGTGTGACAGAGTGTGACAGAGTGTG 1380
QY 1497 GAGCGGAGAGCTTGTGACATGAGAGTGTGACAGAGTGTGACAGAGTGTGACAGAGTGTG 1556
DB 1381 TAGCGGAGAGCTTGTGACATGAGAGTGTGACAGAGTGTGACAGAGTGTGACAGAGTGTG 1440
QY 1557 GCCCAACACCGAGAGAGCCGCAAGTACAGAGCGGAGAGAGCCCTGTGCTCAAGCTTGTG 1616
DB 1441 GCCCAACACCGAGAGAGCCGCAAGTACAGAGCGGAGAGAGCCCTGTGCTCAAGCTTGTG 1500
QY 1617 CTTGTGCGGAGAGGAGAGCCAGCTGACCAAG 1647
DB 1501 TCTGTGCGGAGAGGAGAGCCAGATGACTTAAG 1531

RESULT 11
BC001517 2455 bp mRNA linear PRI 16-SEP-2003
LOCUS Homo sapiens thioesterase, adipose associated, mRNA (cDNA clone
DEFINITION IMAGE:2966784), complete cds.
ACCESSION BC001517
VERSION BC001517.2 GI:33876308
KEYWORDS Homo sapiens (human)
SOURCE
```

ORGANISM

Homo sapiens
Mammalia; Eutelesia; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutelesia; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

1 (bases 1 to 2455)

Strausberg, R., Pelting, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,

Alteich, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshimaru, S.,

Carrinchi, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,

Abrahamson, R.D., Mullaly, S.J., Bosak, S.A., McMan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hilly, S.W.,

Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,

Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E.,

Scherer, A., Schein, J.E., Jones, S.J., and Maitra, M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

2 (bases 1 to 2455)

Strausberg, R.

Direct Submission

Submitted (21-DEC-2000) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

On Aug 19, 2003, this sequence version replaced gi:16306685.

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC/DCTP

cDNA Library Preparation: Rubin Laboratory

DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LIML)

contact: <http://www.systemsbio.org>

contact: amadansystemsbiology.org

Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>

Series: IRAL Plate: 3 Row: k Column: 9

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 25777709

This clone has the following problem: The cds is short compared to

the longest cds in the locus.

Location/Qualifiers

1..2455

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone_image="296784"

/tissue_type="Skin, melanotic melanoma."

/clone_id="N1H MGC_20"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

1..2455

/gene="THEA"

/note="synonyms: KIAA0707, BFIT, STARD14, BFIT1, BFIT2,

THEM1"

/db_xref="LocusID:26027"

/db_xref="MIM:606803"

346..1011

/codon_start=1

/product="THEA protein"

ORIGIN

Query Match 39.4% Score 730.8; DB 9; Length 2455;
Best Local Similarity 97.8%; Pred. No. 3,4e-132;
Matches 741; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

misc_feature

376..807
/note="COG1607; Region: Acyl-CoA hydrolase [lipid
metabolism]"
874..1005
/note="COG1607; Region: Acyl-CoA hydrolase [lipid
metabolism]"
/db_xref="CDD:COG1607"

misc_feature

51 GGGCTTGGCTCTGTGTTCTCCACCGACATCCCGAAGTCAGCTTACGTGGGGGAA 110
276 GGGCTTGGCTCTGTGTTCTCCACCGACATCCCGAAGTCAGCTTACGTGGGGGAA 335
111 CGACAGTCCATGGAGACCGGAGAGGATACCGGAACCCACAGAGCTGACATGACCA 170
336 CGACAGTCCATGGAGACCGGAGAGGATACCGGAACCCACAGAGCTGACATGACCA 395
171 GCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 230
396 GCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 455
231 GTGATTTGACACCAACGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 290
456 GTGATTTGACACCAACGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 515
291 AGCTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 350
516 AGCTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 575
351 CAAGGACCAAGTGAACCGAGGCTTCACTCCAGCATGAGGAGGAGGAGGAGGAGGAGG 410
576 CAAGGACCAAGTGAACCGAGGCTTCACTCCAGCATGAGGAGGAGGAGGAGGAGGAGG 635
411 GAGGACCTGTGCTGTGGAAGACAGTGAATGTGTGAAGGCTTGGCCACCTTGTGGC 470
636 GAGGACCTGTGCTGTGGAAGACAGTGAATGTGTGAAGGCTTGGCCACCTTGTGGC 695
471 CCGCCGAGAGTCAACCAAGGTGAAGCTGAACATCACTGCTGCGGACAGAGAGAGAGAA 530
696 CCGCCGAGAGTCAACCAAGGTGAAGCTGAACATCACTGCTGCGGACAGAGAGAGAGAA 755
531 GATGAGACAGTGTGGGCGCTGAGCGCGCGCATGGCGCTTGTCTATGACAGACCAT 590
756 GATGAGACAGTGTGGGCGCTGAGCGCGCGCATGGCGCTTGTCTATGACAGACCAT 815
591 CAAGGACCTCTTGGCCCAATCTGCGCATTCAGGCGCATTCAGGCGCATTCAGGCGCAT 650
816 CAAGGACCTCTTGGCCCAATCTGCGCATTCAGGCGCATTCAGGCGCATTCAGGCGCAT 875
651 CATGTGCGCGCTGAGAAACCGGTGAGAGAGTGAAGTGTGCTGCTGCTGCTGCTGCTGCT 710
876 CATGTGCGCGCTGAGAAACCGGTGAGAGAGTGAAGTGTGCTGCTGCTGCTGCTGCTGCT 935
711 CAATCACAGGAGCAACCTTTGGGGGCAAGATATGAGCTGAGTGAAGTGAAGTGAAGTGA 770
936 CAATCACAGGAGCAACCTTTGGGGGCAAGATATGAGCTGAGTGAAGTGAAGTGAAGTGA 995
771 CATTCAGACCAAGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 808
996 CATTCAGACCAAGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1033

RESULT 12
 AX870465 751 bp DNA linear PAT 17-DEC-2003
 LOCUS AX870465
 DEFINITION Sequence 5370 from Patent EP1074617.
 ACCESSION AX870465
 VERSION AX870465.1 GI:40025328
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 Oca,T., Isogai,T., Nishikawa,T., Hayaashi,K., Saito,K., Yamamoto,D., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
 Primers for synthesizing full-length cDNA and their use
 Patent: EP 1074617-A 5370 07-FEB-2001;
 Research Association for Biotechnology (ur)

FEATURES
 source location/Qualifiers
 1..751
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 34.3%; Score 636.4; DB 6; Length 751;
 Best Local Similarity 97.8%; Pred. No. 9.6e-114;
 Matches 654; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 19 ATGATCCAGAAATGTCGGAATCACTGCGACGCGGGCTTGCCCTGTGTTCTCCAAACGC 78
 Db 83 ATGATCCAGAAATGTCGGAATCACTGCGACGCGGGCTTGCCCTGTGTTCTCCAAACGC 142

QY 79 ACATCCCGGAAGTCAGCTTAACGTCGCGGGAACGACAGTGCACATGAGACGCGGAGGA 138
 Db 143 ACATCCCGGAAGTCAGCTTAACGTCGCGGGAACGACAGTGCACATGAGACGCGGAGGA 202

QY 139 TACCGGAACCCCAACGAGGTGACAGTATGAGCAGCTGTGCTGCCCTCCACCAACCA 198
 Db 203 TACCGGAACCCCAACGAGGTGACAGTATGAGCAGCTGTGCTGCCCTCCACCAACCA 262

QY 199 CGTGTGAGCTGAGCGTTCGCGGACGCTGCTCAATGATGACACACGCGCTTGCTGTC 258
 Db 263 CGTGTGAGCTGAGCGTTCGCGGACGCTGCTCAATGATGACACACGCGCTTGCTGTC 322

QY 259 GCGGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318
 Db 323 GCGGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 382

QY 319 CACACATTATGTTGGACAAGTGTGAATATCAAGGCCAAGTGAACCGGGCTTCAAC 378
 Db 383 CACACATTATGTTGGACAAGTGTGAATATCAAGGCCAAGTGAACCGGGCTTCAAC 442

QY 379 TCCAGCATGAGAGTGGGATCCAGTGGCCCTCGAGGACCTGTGCTGTGAGAACAGTGG 438
 Db 443 TCCAGCATGAGAGTGGGATCCAGTGGCCCTCGAGGACCTGTGCTGTGAGAACAGTGG 502

QY 439 AATGTGTGCAAGGCTTGGCCACCTCTGTGCGCCCGCGAGATCAACCAAGTGAAGCTG 498
 Db 503 AATGTGTGCAAGGCTTGGCCACCTCTGTGCGCCCGCGAGATCAACCAAGTGAAGCTG 562

QY 499 AAGCAGATCAAGCGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 558
 Db 563 AAGCAGATCAAGCGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 622

QY 559 CGGCGCATGCGCTTGTCTATGACACCATCAAGGACCTTCTGGCCACTGCGCCATT 618
 Db 623 CGGCGCATGCGCTTGTCTATGACACCATCAAGGACCTTCTGGCCACTGCGCCATT 682

QY 619 CAGGCGCATCTGG-AGAGCAGAGACTGAGCGCGCATGTCGCGGCTGAGAGACCCGCT 677
 Db 683 CAGGCGCATCTGGAAAAAGCAGAGACTGAGCGCGCATGTCGCGGCTGAGAGACCCGCT 742

QY 678 CGAGACTGT 686
 Db 743 NGAAATGT 751

RESULT 13
 BD150527 751 bp DNA linear PAT 17-JAN-2003
 LOCUS BD150527
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD150527
 VERSION BD150527.1 GI:27856285
 KEYWORDS JP 2002191363-A/5370.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 751)
 Oca,T., Isogai,T., Nishikawa,T., Hayaashi,K., Saito,K., Yamamoto,D., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
 Primers for synthesizing full-length cDNA and use thereof
 Patent: JP 2002191363-A 5370 09-JUN-2002;
 HELIX RESEARCH INSTITUTE

FEATURES
 source location/Qualifiers
 1..751
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 34.3%; Score 636.4; DB 6; Length 751;
 Best Local Similarity 97.8%; Pred. No. 9.6e-114;
 Matches 654; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 19 ATGATCCAGAAATGTCGGAATCACTGCGACGCGGGCTTGCCCTGTGTTCTCCAAACGC 78
 Db 83 ATGATCCAGAAATGTCGGAATCACTGCGACGCGGGCTTGCCCTGTGTTCTCCAAACGC 142

QY 79 ACATCCCGGAAGTCAGCTTAACGTCGCGGGAACGACAGTGCACATGAGACGCGGAGGA 138
 Db 143 ACATCCCGGAAGTCAGCTTAACGTCGCGGGAACGACAGTGCACATGAGACGCGGAGGA 202

QY 139 TACCGGAACCCCAACGAGGTGACAGTATGAGCAGCTGTGCTGCCCTCCACCAACCA 198
 Db 203 TACCGGAACCCCAACGAGGTGACAGTATGAGCAGCTGTGCTGCCCTCCACCAACCA 262

QY 199 CGTGTGAGCTGAGCGTTCGCGGACGCTGCTCAATGATGACACACGCGCTTGCTGTC 258
 Db 263 CGTGTGAGCTGAGCGTTCGCGGACGCTGCTCAATGATGACACACGCGCTTGCTGTC 322

QY 259 GCGGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318
 Db 323 GCGGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 382

QY 319 CACACATTATGTTGGACAAGTGTGAATATCAAGGCCAAGTGAACCGGGCTTCAAC 378
 Db 383 CACACATTATGTTGGACAAGTGTGAATATCAAGGCCAAGTGAACCGGGCTTCAAC 442

QY	379	TCACACATAGAGATGGGCATTCACAGGGGCGCTGGAGGACCTTGCTCTTGAAGAAGCATGG	438
Db	443	TCCACACATGGAGTGGGACATCCAGTGGCGCTGGAGGACCTTGCTCTTGAAGAAGCATGG	502
QY	439	AATGTGTGCAGGCGCTTGGCCACTTGTGGCCCGCCGAGATCACCAAGGTGAAGCTG	498
Db	503	AATGTGTCCAAAGGCGCTTGGCCACTTGTGGCCCGCCGAGATCACCAAGGTGAAGCTG	562
QY	499	AAGCAGATCACCGCCCGGACAGAAAGAGAGATGAGACACAGTGTGGCGGTGAGCGC	558
Db	563	AAGCAGATCACCGCCCGGACAGAAAGAGAGATGAGACACAGTGTGGCGGTGAGCGC	622
QY	559	CGGCGCATAGCGCTTGTCTATGCACAGACACCATCAAGAGACTCTCTGGCCAACTGGCGCATT	618
Db	623	CGGCGCATAGCGCTTGTCTATGCACAGACACCATCAAGAGACTCTCTGGCCAACTGGCGCATT	682
QY	619	CAGGCGCATCTGG-AGAGCAGAGACTGTAGCCGCATGTGTCCGGCTGAGAGAACCCGATG	677
Db	683	CAAGCGCATCTGGAAAAGACANMAGACTGTAGCCGCATGTGTCCGGCTTAAATAAACCTGT	742
QY	678	GGAGAGTGT 686	
Db	743	NGAATATGT 751	

RESULT	14
LOCUS	AB078618
DEFINITION	Mus musculus mCACH-1 mRNA for cytosolic acetyl-CoA hydrolase,
ACCESSION	complete cds.
VERSION	AB078618
KEYWORDS	AB078618.1 GI:18307691
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 Suematsu,N., Okamoto,K. and Isohashi,F. cytosolic acetyl-CoA hydrolase Published Only in Database (2002) 2 (bases 1 to 1689) Suematsu,N., Okamoto,K. and Isohashi,F. Direct Submission Submitted (22-JAN-2002) Naoya Suematsu, St Marianna University School of Medicine, Department of Biochemistry, Miyamae-ku, Sugao 2-16-1, Kawasaki, Kanagawa 216-8511, Japan (E-mail:n2susue@marianna-u.ac.jp), Tel:81-44-977-8111(ex.3525) , Fax:81-44-976-7553)
AUTHORS	JOURNAL
TITLE	TITLE
REFERENCE	JOURNAL
AUTHORS	TITLE
TITLE	JOURNAL

FEATURES
SOURCE

gene
CDS

```

1. tissue_type="liver"
/ctissue_type="liver"
1.1689
/ gene="mCACH-1"
8.1678
/ gene="mCACH-1"
/codon_start=1
/product="cytosolic acetyl-CoA hydrolase"
/protein_id="BAB84021.1"
/db_xref="GI:18307692"
/translation="MMSWAPGPGVLSQALQPAHADSREGSLSGQLLNKWDPTACLAAL
EKHAGISCTYASMDLILFDLTARIGQITIRAKVTAFSTSMELSIKIVLYQDFRTGIQ
KILCVAFSTFVKAPEKKEVYH.KPVLLQTEQEVNHLNLSERKRVLRQHNFTNNIKML
BSRRSDRLCSNKEEAGATMTGISVOSIELVLPQEHANHGNTPGQILNANMEVATLISA
SRLGHGHPILKSVDMKEKPGSGVGRVLRVLSAFLVNNTPQNSVEVGRVEAFDQCEVAB
CGGRINLSAOWDILYNNAVDDDEKLTTPPRITQISKQDIDRRVQGLARRRIRIGRRVISH
KEEVLNLSQWDLISQNSLNTVNAALKNLASKGWEITTTTLEIKITLYLEQDAISVYK

```

ORIGIN

Query Match	23.4%	Score 434.8;	DB 10;	Length 1689;
Best Local Similarity	57.9%;	Pred. No. 1.6e-74;		
Matches 854;	Conservative	0;	Mismatches 607;	Indels 15; Gaps 4;

QY	154	GAGGTGCAGATGAGCAGCACTGGTGTGGCCCTGGCCACACCAACCAACGATGTGAGCTGAGC	213
Db	32	GAGGTGCTCATGAGCCAGGAGCCATCCAGCCGGGTCTATGCCAATCTCCCGGCGAGCTGAGC	91
QY	214	GTGGGGCACTGCTCAAGTGTGATTTGACACCAACGAGTCTCTGTCCGCGAGAGGAGCAAGCT	273
Db	92	GCAGGGCAGCTGTCTCAAGTGTGATGAGACCAACCGCTTCTGCGGGCGGAAAGATGCT	151
QY	274	GAGTGGCCCTGTGTCAACGCTTCCATGGATGACATCTATTTTGTAGACACCAATTAGTGT	333
Db	152	GAGATTTCTGTGTGACAGCCCTCCATGATGATACATTTGTTTGTAGGACACAGCCGAAATT	211
QY	334	GGAACAAGGTGGAATATTAAGGCCAAGGTGAACCGGGGCTTCAACTCAGATGTGAGTGT	393
Db	212	GGAACAATTTATTCATCATGAGACAAATAATGACTATGAGGGTTTCAGACACAAGATGTGAGTCT	271
QY	394	GGCATCCAGGTGAGCTCGAGAGACCTGTGTCTGTGAGAAGCAGTGGAAATGTGTCAAGGCC	453
Db	272	AGTATCAAGGTCAATAGTCAAGACAAAGTTCAACGGGATCCAGAAAGTCTCTGTGCTGCT	331
QY	454	TTGGCCACTTTCGTGGCCCGCCGAGAGATCA--CCAAGTGAAGCTGAACAGATCAAG	510
Db	332	TTCTCTACGTTGTGTAGCTTAAACAGTGTGGCAAAAGAAAGTTCACTTAAACTGTCTGT	391
QY	511	CCGGGACAGAGAAGAGGAGAAATGAGACAGTGTGGGCTGAGCGCCGCGCATGTGC	570
Db	392	CTTCAACACAGACAGAAACAAGTGAAGACATCTGGCTTATAGAGAGAAAGAAAGTCCGA	451
QY	571	CTTGTCTATGTGAGACACCATCAAGGACCTCTGTGGCCAACTCGGCATTTCAAGGCGATCTG	630
Db	452	CTGACAGATGAGAACACTTTCACAACTTTATGAAGAGAGACAGATTTAGCGATTTCC	511
QY	631	GAGAGCAAGACCTGTAGCCCGCATGTGTCCGGCTGAGAAAGACCCGTGTGAGAGTGTGAG	690
Db	512	ATTGTATGATGAAGAAGAAAGAACCGGCACCACTATGGGACCTCTGTCAAGACATCGAG	571
QY	691	CTGGTCTCTGCTCCCGCCAGCGCAATCACAGGGCAACACTTTGGGGGCGCAATATATAGCC	750
Db	572	CTTGTCTCTCCGCCCCACAGCAAACTCATACGAAACAACATTTGTGTGGCGAATATATGCA	631
QY	751	TGATGTGAGATGTGGCCACCATTTGCAGCCAGCGGCTCTGCGCCACTTACAGCTG	810
Db	632	TGATGTGAGACAGTCCGCAACATTTCTGTCAAGCCGCGCTGTGTCAATGGGCATTCCTTTCTG	691
QY	811	AAGGCCATTTGAATGTTTCCACTTTCGAGGCCCGTCCAGGTGGCGAACCGTCTGTGTCTC	870
Db	692	AAGTCTGTGATATGTTTAAATTCGGGGAGCCGTCCACAGTTTGAGACCGGCTTGTCTTC	751
QY	871	AAABCCATTCGTGAACATATGCTTCAAACTATGACATGAGAGTGGGCGTGTGTGTGAGAGC	929
Db	752	AGTGCACATAGTCAACAACACCTTTCACAACAAGTGTGAAAGTGCAGTGCGTGTGTGAAGCC	811
QY	930	-----CTATGCGCCAGAGGCTGAGACCCACCGGCGGCACATCAACATGATGCTTATGATCC	984
Db	812	TTTGTACTGTCAAGAGTGGGCGAGAGGCGCAAGGCCGCAATCAACAGCGCTTTTCTCATTT	871
QY	985	TTTGTGTCTGTGACGAGATGACCAAGCCCAATGTGTGCTGTGATTTGGGCCCAAGCCC	1044
Db	872	TACAATGCTGTGTATGTACAGAGAAATCTATCACTTTTCCAGAAATCCAACCATTTCA	931
QY	1045	GCGCATGTGTAGCGCGGCTTACCGAGAGGCCATGTGCCAAGAAAGAAATTCGCTGTGACAGG	1104
Db	932	AAGGACATATTTGCGCCTTATCCAGGAGGCATGTCAACGAGAGAAATTCGCTGTGAGAGG	991

OY	865	GTCGCTCAAGCCATCGTGAACAATACCTTTCAAACGTAACTGAGAGTGGGGTGGCGCTG	924
Db	759	GCTTTCATTGCCATGCTGTCAACAACCTTTTCAAGAACGTGTGAAGTGTGGCGTAGCTGT	818
OY	925	GAGGC-----CTATGCGCAGGAGGCTGAGACCACCGGCGCCCATCATCAAGTGTGCTTT	978
Db	819	GAGGCGCTTGAAGTGTGCGGAGTGGGCCGAGGGCCGAAGGCCGCAATCAACAGCGCTTTT	878
OY	979	ATGACCTTTGTGTCTTGGAGCGCAGATGACCAAGCCCCAAGTTGCTGCTCCGTGATTCGSCCC	1038
Db	879	CTCATTTTCAATGCTGTGTGATGACCAAGAGGAACCTCATCACCTTTCCAGAATCCAGCCC	938
OY	1039	CAGCCCCGAGATGCTGTGAGCGGGCGGTACCGAGAGCGCATGCGCAAGAAACAATCCGCTG	1098
Db	939	ATTTTCAAGAGTATTTTTTCAGACGATACAGAGGACTATTGCAAGGAGAAATTCGCTTG	998
OY	1099	GACAGGAAGTAATCTGTGCTCTGTAAAGCAGACAGAGTGCCCTCTCCGTCCCTCGGAC	1158
Db	999	GGCAGGAATAATGTATTATTTCC---CACAAAGAAAGTCCCATCTTGGCAGCANATGGAT	1055
OY	1159	CCTAGCAACCAAGGTGTACCTGAGCTCAATACGTCTTCTCTTGAAGATGCTTGTGGCC	1218
Db	1056	ATPAGCAAAAAGGATCCATTAAGTAACCAACATGTGAAGCTCTCAAAAAATCTGGCATTC	1115
OY	1219	AAGGACAACGTGGTGCTGTCTCTCGAGATTCAGTCAAGTTCGCCCTGTACCTCTGAGGAT	1278
Db	1116	AAAAACGGTTGGAGATATCAACAACCTTGAGAGAAATAAATATATACTGGAAGG	1175
OY	1279	GACAAGTTCCTCTCTCTTCACATGAGAGATGTGTGCAATGTGAATGCAAGCCAGGACCTTC	1338
Db	1176	CAGATGCGCATATCTGTAAAGGTGGAAAAACAGGTGTGAANTCCAGCCCGCGTGGCTTAT	1235
OY	1339	CTGCTGCTCTTGGACCTGCGTCAAGGCGCAGATGGGACAAGCACTAACCGAGCGTGGAG	1398
Db	1236	CATCTCTGTGTCGACTTCACAAAAGCGCCCTTTATGGGACCCCCATPACATATCTTTGAA	1295
OY	1399	CTAGTGCAGCGGTAAACGAGSAGAGAGCGCATCTAACACGTCACACAGCCCTGCGCCTGGA	1458
Db	1296	GTCATATGACCAAGGTGAGTGAAGAGATCAAGTATATTAATCAATCACTGTGTGTGTAAAT	1355
OY	1459	GGTCACACAAAGCCCCCAGAGCTTGATGATCTCTGCGCTCGAAGCGGAAGCCTTGACAAAT	1518
Db	1356	GGAGACAAA---CCCAAGGACTTGTGTGTAATCTGTGTCACAAAAGAAAGCCCTTAAGAC	1412
OY	1519	GGGGAACCCCTATGTCAATCGCGCTGAGGTGCGTCAAGCTGCCACACACCAAGAGACGCCA	1578
Db	1413	GACAAACCTAATCGTGGCAGCTATATGCTAGTGTACTGCGGCTGTGCCCTCATCTCCA	1472
OY	1579	GAGTACAGAGCGCGGAGAGACCCCTGCTCAAGGCTTGTGCTGCTTGCCTTGGC	1624
Db	1473	CAGTACATCAAGAACCGACAGTATTTGTGTGGGTTCTCATTCACGC	1518

> O < IntellGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file us-10-055-624b-2.res made by tport on Tue 24 Aug 104 10:14:57-PST.

Query sequence being compared: US-10-055-624B-2 (1-607)
Number of sequences searched: 3
Number of scores above cutoff: 3

Results of the initial comparison of US-10-055-624B-2 (1-607) with:
File : US10055624B.pep

```

100-
N -
U -
M -
B -
E -
R -
O -
F -
S -
E -
Q -
U -
N -
C -
E -
S -
SCORE 0 67 135 202 270 337 405 472 540 607
STDEV -2 -1 -1 -1 -1 -1 -1 -1 -1 -1

```

PARAMETERS

Similarity matrix PAM-150 K-tuple 1
Threshold level of sim. 16%
Mismatch penalty 1 Joining penalty 20
Gap penalty 5.00 Window size 500
Gap size penalty 0.05
Cutoff score 1
Randomization group 0

SEARCH STATISTICS

Scores: Mean 496 Median 327 Standard Deviation 149.90
Times: CPU 00:00:00.00 Total Elapsed 00:00:00.00

Number of residues: 1795
Number of sequences searched: 3
Number of scores above cutoff: 3

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

Sequence Name	Description	Init. Opt.	Sig. Frame
		Length Score	Score

1. US-10-055-624B-2 Sequence 2, Application U 607 607 607 0.74 0
The list of other best scores is:

Sequence Name	Description	Length Score	Init. Opt.	Sig. Frame
2. US-10-055-624B-4 Sequence 4, Application U	594 557 563 0.41 0			
3. US-10-055-624B-6 Sequence 6, Application U	594 326 529 -1.13 0			

1. US-10-055-624B-2 (1-607)
US-10-055-624B-2 Sequence 2, Application US/10055624B

Initial Score = 607 Optimized Score = 607 Significance = 0.74
Residue Identity = 100% Matches = 607 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

```

X 10 20 30 40 50 60 70
MIQVGNHRLRGGLASVFSNRTSRKSKALRAGNDSAMADGGYNPTEVQNSQVLPCHTNOGRLSVGOLKX
MIQVGNHRLRGGLASVFSNRTSRKSKALRAGNDSAMADGGYNPTEVQNSQVLPCHTNOGRLSVGOLKX
X 10 20 30 40 50 60 70

```

```

80 90 100 110 120 130 140
IDTTACTSARHAGCCPVTASMDIYFEHTISVGQVNIKAKVNRFPNSMEVGIOVASDLCSEKQWVCK
IDTTACTSARHAGCCPVTASMDIYFEHTISVGQVNIKAKVNRFPNSMEVGIOVASDLCSEKQWVCK
80 90 100 110 120 130 140

```

```

150 160 170 180 190 200 210
ALATFVARREITVKKQITPRTEEEKMEHSVAERRRMELVYADITKDLANCAIQGLDESRCRMPAE
ALATFVARREITVKKQITPRTEEEKMEHSVAERRRMELVYADITKDLANCAIQGLDESRCRMPAE
150 160 170 180 190 200 210

```

```

220 230 240 250 260 270 280
KTRVESVELVLPNHNQNTFGQIMAMMENVAITIASRLCAHPTLKAIEHFHRRGSPQVGRVLYKAIY
KTRVESVELVLPNHNQNTFGQIMAMMENVAITIASRLCAHPTLKAIEHFHRRGSPQVGRVLYKAIY
220 230 240 250 260 270 280

```

```

290 300 310 320 330 340 350 360
NNAFKHSMEEVGCVEAYROEAEHRRHINSAPMTFVLDADDDQQLIPWIRPOGDERRRYREASARKIRL
NNAFKHSMEEVGCVEAYROEAEHRRHINSAPMTFVLDADDDQQLIPWIRPOGDERRRYREASARKIRL
290 300 310 320 330 340 350 360

```

```

370 380 390 400 410 420 430
DRKTIYVCKQTEVPLSVPMDSNOVYLSYNNVSLKMLVAKDNVLSSEISQVRLYTLDDKFLSFHMEVY
DRKTIYVCKQTEVPLSVPMDSNOVYLSYNNVSLKMLVAKDNVLSSEISQVRLYTLDDKFLSFHMEVY
370 380 390 400 410 420 430

```

```

440 450 460 470 480 490 500
HVDAAQAFLLISDLRORPEMDKHYSVELVQVDEDDAIYHTVSPALGHTKPPDFTYILASRRKPCNDGPY
HVDAAQAFLLISDLRORPEMDKHYSVELVQVDEDDAIYHTVSPALGHTKPPDFTYILASRRKPCNDGPY
440 450 460 470 480 490 500

```

```

510 520 530 540 550 560 570
VIALRSVTLPTHTRETPPYRGETLCSGFCIMREDDQLTKCWVWVSLTELYSAGSFYWGLESRSKGRSDG
VIALRSVTLPTHTRETPPYRGETLCSGFCIMREDDQLTKCWVWVSLTELYSAGSFYWGLESRSKGRSDG
510 520 530 540 550 560 570

```

```

580 590 600
WNGKLAGHLSLTKAIIPVAKINSRFGYLQDT X
WNGKLAGHLSLTKAIIPVAKINSRFGYLQDT X
580 590 600

```

2. US-10-055-624B-2 (1-607)

US-10-055-624B-4 Sequence 4, Application US/10055624B

Residue Identity	=	557	Optimized score	=	563	Significance	=	0.41
Gaps	=	0	Matches	=	547	Mismatches	=	.46
			Conservative Substitutions					1

X	10	20	30	40	50	60	70																																																						
M	O	N	G	N	H	R	R	G	L	A	V	S	F	S	N	R	S	R	K	S	A	L	A	R	G	N	D	S	A	M	A	D	G	E	R	N	T	E	O	S	V	L	P	C	H	T	N	O	R	G	L	S	T	G	L	L	K				
X	10	20	30	40	50	60	70																																																						
M	O	N	G	N	H	R	R	G	L	A	V	S	F	S	N	R	S	R	K	S	A	L	A	R	G	N	D	S	A	M	A	D	G	E	R	N	T	E	O	S	V	L	P	C	H	T	N	O	R	G	L	S	T	G	L	L	K				
X	10	20	30	40	50	60	70																																																						
M	O	N	G	N	H	R	R	G	L	A	V	S	F	S	N	R	S	R	K	S	A	L	A	R	G	N	D	S	A	M	A	D	G	E	R	N	T	E	O	S	V	L	P	C	H	T	N	O	R	G	L	S	T	G	L	L	K				
X	10	20	30	40	50	60	70																																																						
I	D	T	T	A	C	L	S	E	R	H	A	G	C	P	C	T	A	S	M	D	I	F	E	H	T	I	S	V	G	O	V	N	I	A	K	N	R	A	F	S	N	M	E	V	I	O	A	S	E	D	I	C	S	E	K	O	N	A	V	C	K
X	10	20	30	40	50	60	70																																																						
I	D	T	T	A	C	L	S	E	R	H	A	G	C	P	C	T	A	S	M	D	I	F	E	H	T	I	S	V	G	O	V	N	I	A	K	N	R	A	F	S	N	M	E	V	I	O	A	S	E	D	I	C	S	E	K	O	N	A	V	C	K
X	10	20	30	40	50	60	70																																																						
A	L	A	T	V	A	R	R	E	I	T	K	Y	K	L	O	I	T	R	E	E	K	M	E	S	V	A	E	R	R	M	L	V	A	D	I	T	K	O	L	L	A	N	C	A	I	O	G	D	L	E	S	R	D	S	R	N	V	P	A	R	
X	10	20	30	40	50	60	70																																																						
A	L	A	T	V	A	R	E	I	T	K	Y	K	L	O	I	T	R	E	E	K	M	E	S	V	A	E	R	R	M	L	V	A	D	I	T	K	O	L	L	A	N	C	A	I	O	G	D	L	E	S	R	D	S	R	N	V	P	A	R		
X	10	20	30	40	50	60	70																																																						
K	I	R	V	A	S	S	E	L	V	L	P	P	H	A	N	I	O	G	N	F	G	O	I	A	M	E	N	V	A	T	I	A	S	L	C	A	H	P	T	L	K	A	I	M	F	H	R	G	S	O	G	D	L	V	I	K	A	I	V		
X	10	20	30	40	50	60	70																																																						
K	I	R	V	A	S	S	E	L	V	L	P	P	H	A	N	I	O	G	N	F	G	O	I	A	M	E	N	V	A	T	I	A	S	L	C	A	H	P	T	L	K	A	I	M	F	H	R	G	S	O	G	D	L	V	I	K	A	I	V		
X	10	20	30	40	50	60	70																																																						
N	N	A	F	H	S	M	E	V	C	V	E	A	R	O	E	A	T	H	R	R	H	I	N	S	A	F	M	T	F	V	L	D	A	D	O	P	O	L	L	P	I	R	O	P	D	G	E	R	R	A	R	K	I	R	L						
X	10	20	30	40	50	60	70																																																						
N	N	A	F	H	S	M	E	V	C	V	E	A	R	O	E	A	T	H	R	H	I	N	S	A	F	M	T	F	V	L	D	A	D	O	P	O	L	L	P	I	R	O	P	D	G	E	R	R	A	R	K	I	R	L							
X	10	20	30	40	50	60	70																																																						
D	R	K	I	V	S	C	K	O	T																																																				

[illegible]

> O <
O | 0 IntellGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file us-10-055-624b-4.res made by tport on Tue Aug 24 Aug 104 10:16:24-PST.

Query sequence being compared: US-10-055-624B-4 (1-594)
Number of sequences searched: 3
Number of scores above cutoff: 3

Results of the initial comparison of US-10-055-624B-4 (1-594) with:
File : US10055624B.pep

```

100-
-
N
-
U 50-
-
M
-
B
-
E
-
R
-
O
-
F 10-
-
S
-
B 5-
-
U
-
O
-
N
-
E
-
C
-
S
-
0-----
SCORE 0 66 132 198 264 330 396 462 528 594
STDEV -2 -1 -1 * *

```

PARAMETERS

Similarity matrix PAM-150 K-tuple 1
Threshold level of sim. 16%
Mismatch penalty 1 Joining penalty 20
Gap penalty 5.00 Window size 500
Gap size penalty 0.05
Cutoff score 1
Randomization group 0

SEARCH STATISTICS

Scores: Mean 491 Median 323 Standard Deviation 147.97
Times: CPU 00:00:00.00 Total Elapsed 00:00:00.00

Number of residues: 1795
Number of sequences searched: 3
Number of scores above cutoff: 3

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

Sequence Name	Description	Init. Opt.	Length	Score	Sig.	Frame

The list of other best scores is:

1. US-10-055-624B-4 Sequence 4, Application U 594 594 594 0.70 0

Sequence Name	Description	Length	Score	Sig.	Frame
2. US-10-055-624B-2 Sequence 2, Application U	607 559 565 0.46 0				
3. US-10-055-624B-6 Sequence 6, Application U	594 322 558 -1.14 0				

1. US-10-055-624B-4 (1-594)

US-10-055-624B-4 Sequence 4, Application US/10055624B

Initial Score = 594 Optimized Score = 594 Significance = 0.70
Residue Identity = 100% Matches = 594 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

```

X 10 20 30 40 50 60 70
MIQVGNHRLRGGLASVFSNRTSRKSLRAGNDSAMADGEGYRNPTEVMSQVLVPCHTNQRGLSVGLKX
X 10 20 30 40 50 60 70
X 10 20 30 40 50 60 70

```

```

80 90 100 110 120 130 140
IDTTACLSAERHAGCCPVASMDIYFEHTTISVGQVNIKAKVNRAFNSMEVGIQVASEDLCEKQWVCK
IDTTACLSAERHAGCCPVASMDIYFEHTTISVGQVNIKAKVNRAFNSMEVGIQVASEDLCEKQWVCK
IDTTACLSAERHAGCCPVASMDIYFEHTTISVGQVNIKAKVNRAFNSMEVGIQVASEDLCEKQWVCK

```

```

150 160 170 180 190 200 210
ALATFVARRRITKVKLKQITPRTEEEKMEHSVAERRRMRLVYADITKOLLNACALQGLDLSRDCSMVPAE
ALATFVARRRITKVKLKQITPRTEEEKMEHSVAERRRMRLVYADITKOLLNACALQGLDLSRDCSMVPAE
ALATFVARRRITKVKLKQITPRTEEEKMEHSVAERRRMRLVYADITKOLLNACALQGLDLSRDCSMVPAE

```

```

220 230 240 250 260 270 280
KTRVESVEVLPPHANHQNTFGQIMAMMENVATTASRLCAHPTLKAIEHFHFRGSPQVGRVLVKAIV
KTRVESVEVLPPHANHQNTFGQIMAMMENVATTASRLCAHPTLKAIEHFHFRGSPQVGRVLVKAIV
KTRVESVEVLPPHANHQNTFGQIMAMMENVATTASRLCAHPTLKAIEHFHFRGSPQVGRVLVKAIV

```

```

290 300 310 320 330 340 350 360
NNAFKSMEEYGVCEAVROBAETHRRHINSAPMTFVVLDDDDQPLLPMTRPQDGERERYREASARKIRL
NNAFKSMEEYGVCEAVROBAETHRRHINSAPMTFVVLDDDDQPLLPMTRPQDGERERYREASARKIRL
NNAFKSMEEYGVCEAVROBAETHRRHINSAPMTFVVLDDDDQPLLPMTRPQDGERERYREASARKIRL

```

```

370 380 390 400 410 420 430
DRKIYVCKQTEVEPLSVPMPSNQVYLSYNNVSLKMLVAKDMVLSSEISOVRLYTLLEDDKLSFHEWV
DRKIYVCKQTEVEPLSVPMPSNQVYLSYNNVSLKMLVAKDMVLSSEISOVRLYTLLEDDKLSFHEWV
DRKIYVCKQTEVEPLSVPMPSNQVYLSYNNVSLKMLVAKDMVLSSEISOVRLYTLLEDDKLSFHEWV

```

```

440 450 460 470 480 490 500
HVDAAQAFLLSLDLRQRPMDKHYRSVELVQVDEDDAIHVTSFALGHTKPDFFYLASRRKPCNGDPY
HVDAAQAFLLSLDLRQRPMDKHYRSVELVQVDEDDAIHVTSFALGHTKPDFFYLASRRKPCNGDPY
HVDAAQAFLLSLDLRQRPMDKHYRSVELVQVDEDDAIHVTSFALGHTKPDFFYLASRRKPCNGDPY

```

```

510 520 530 540 550 560 570
VIALRSVTLPTHRETPYRGETLCSGFCILMRBDQLTKVSYNQATPPGVNLVYTVNAGLSSFFYTFKAC
VIALRSVTLPTHRETPYRGETLCSGFCILMRBDQLTKVSYNQATPPGVNLVYTVNAGLSSFFYTFKAC
VIALRSVTLPTHRETPYRGETLCSGFCILMRBDQLTKVSYNQATPPGVNLVYTVNAGLSSFFYTFKAC

```

```

580 590 X
EQFLDNRNDLAPSLQTL
EQFLDNRNDLAPSLQTL
EQFLDNRNDLAPSLQTL

```


IntelliGenetics

FastDB - Fast Pairwise Comparison of Sequences

Release 5.4

Results file us-10-055-624b-1.res made by tport on Tue 24 Aug 104 10:18:19-PST.

Query sequence being compared: US-10-055-624B-1 (1-1857)
Number of sequences compared: 30

Number of sequences searched:	20
Number of scores above cutoff:	20

Results of the initial comparison of US-10-055-624B-1 (1-1857) with:
File : US10055624B.seq

100 -
N -
U 50 -
M -
R -
E -
R * -
O -
F 10 -
S -
R 5 -
O -
C -
U -
E -
N -
C -
R -
S -

SCORE 0 206 413 619 825 1032 1238 1444 1651 1857

STDEV 1 2 3

* * *

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	5.00	Window size	500
Gap size penalty	0.33		
Cutoff score	1		
Randomization group	0		

SEARCH STATISTICS

	Mean	Median	Standard Deviation
Scores:	231	16	553.73
Times:	CPU		Total Elapsed
	00:00:00.00		00:00:00.00

Number of residues:	6764
Number of sequences searched:	20
Number of scores above cutoff:	20

The scores below are sorted by initial score. The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

Sequence Name	Description	Length	Init. Opt. Score	Frame

1. US-10-055-624B-1 Sequence 1, Application U	1857	1857	2.94	0
---	------	------	------	---

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
---------------	-------------	--------	-------------	------------	------	-------

	****	2 standard deviations above mean	****	
2. US-10-055-624B-3 Sequence 3, Application U	1818	1633	1692	2.53
**** 1 standard deviation above mean ****				
3. US-10-055-624B-5 Sequence 5, Application U	2699	884	1453	1.18

1. US-10-055-624B-1 (1-1857)

Initial Score	=	1857	Optimized Score	=	1857	Significance	=	2..94
Residue Identity	=	100%	Matches	=	1857	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=			=	0

	10	20	30	40	50	60	70
X	10	20	30	40	50	60	70
GTGGAATG	CCCTTCAAA	ATCGAATG	TCGAAA	TCACTG	CGACGG	CGCTTGT	GTCTCT
GTGGAATG	CCCTTCAAA	ATCGAATG	TCGAAA	TCACTG	CGACGG	CGCTTGT	GTCTCT
GTGGAATG	CCCTTCAAA	ATCGAATG	TCGAAA	TCACTG	CGACGG	CGCTTGT	GTCTCT
X	10	20	30	40	50	60	70

80 90 100 110 120 130 140
 AACCGCATATCCGGAAAGTCAAGCTTACGTGCGGGGAAACACAGTCCATGCGACAGCGCGAGGAGTACCGG
 AACCGCATATCCGGAAAGTCAAGCTTACGTGCGGGGAAACACAGTCCATGCGACAGCGCGAGGAGTACCGG
 80 90 100 110 120 130 140

150 160 170 180 190 200 210
AACCACGAGGTCAGATGACACCTGTGTGCTCCACACCAACCAAGTGTAGCTAGCGT
AACCACGAGGTCAGATGACACCTGTGTGCTCCACACCAACCAAGTGTAGCTAGCGT
AACCACGAGGTCAGATGACACCTGTGTGCTCCACACCAACCAAGTGTAGCTAGCGT
AACCACGAGGTCAGATGACACCTGTGTGCTCCACACCAACCAAGTGTAGCTAGCGT
150 160 170 180 190 200 210

220 230 240 250 260 270 280
GGGAGCTGCTCAAGTSGATTGAACAACAGGCTTGCCGCGGAGAGGACGCTGGCTGCCCCGTGGT
GGGAGCTGCTCAAGTSGATTGAACAACAGGCTTGCCGCGGAGAGGACGCTGGCTGCCCCGTGGT
GGGAGCTGCTCAAGTSGATTGAACAACAGGCTTGCCGCGGAGAGGACGCTGGCTGCCCCGTGGT
220 230 240 250 260 270 280

290 300 310 320 330 340 350 360
AAGCTTCATGATGACATCTATTGAGCACCCATTGCTGTGGACAAGTGGATATCAAGCCAA
AAGCTTCATGATGACATCTATTGAGCACCCATTGCTGTGGACAAGTGGATATCAAGCCAA
AAGCTTCATGATGACATCTATTGAGCACCCATTGCTGTGGACAAGTGGATATCAAGCCAA
290 300 310 320 330 340 350 360

370 380 390 400 410 420 430
GTGAACCGGGCTTTAACTCCAGCATGAGAGTGGGACATCAAGGAGGCTTGGAGGACCTGTGCTTGAA
GTGAACCGGGCTTTAACTCCAGCATGAGAGTGGGACATCAAGGAGGCTTGGAGGACCTGTGCTTGAA
GTGAACCGGGCTTTAACTCCAGCATGAGAGTGGGACATCAAGGAGGCTTGGAGGACCTGTGCTTGAA

[illegible]

ATCACCCGCGGACGAGAGAGAGATGGAGCACAAGTGTGGCGGCTGAGCCGCGCATGGCCTTGT
510 520 530 540 550 560 570

580 590 600 610 620 630 640
TATGCAACCATCAAGACCTCTGCGCAATCGCATTTGAGAGCAGACTGTAG
TATGCAACCATCAAGACCTCTGCGCAATCGCATTTGAGAGCAGACTGTAG
TATGCAACCATCAAGACCTCTGCGCAATCGCATTTGAGAGCAGACTGTAG
580 590 600 610 620 630 640

1520	1530	1540	1550	1560	1570	1580
GACATATGGGAGACCTTATATGTCATCGCGTGAAGTGTGATACCGTACGCCACACGACGATGACGCCAGAGTA						
GACATATGGGAGACCTTATATGTCATCGCGTGAAGTGTGATACCGTACGCCACACGACGATGACGCCAGAGTA						
1590	1600	1610	1620	1630	1640	1650
AGACGCGGAGAGACCCCTGTGCTCAGGCTTCTTGCCTCTGCGCGAGAGGGGACGACGTGACCAAGTCTGTGCG						
AGACGCGGAGAGACCCCTGTGCTCAGGCTTCTTGCCTCTGCGCGAGAGGGGACGACGTGACCAAGTCTGTGCG						
1660	1670	1680	1690	1700	1710	1720
GTTAGGCTTCTCCTACTAGCTAGCTGATCTTCGCGACATGCGCTTCTATTTCTGTGGGCTCGAATCTCAGTCAAA						
GTTAGGCTTCTCCTACTAGCTAGCTGATCTTCGCGACATGCGCTTCTATTTCTGTGGGCTCGAATCTCAGTCAAA						
1730	1740	1750	1760	1770	1780	1790
GCTGCGAGAGACGACGCTTGGAAATGGAAACTAGCTGAGAGACACTGTAAGTACTTTAAAGAAATCCCGCTG						
GCTGCGAGAGACGACGCTTGGAAATGGAAACTAGCTGAGAGACACTGTAAGTACTTTAAAGAAATCCCGCTG						
1800	1810	1820	1830	1840	1850	1860
GCCAAATTCACACGCCATTTGGATATGATACCTTTAAAGACACTTAACCTTATATCAGC						
GCCAAATTCACACGCCATTTGGATATGATACCTTTAAAGACACTTAACCTTATATCAGC						
1870	1880	1890	1900	1910	1920	1930
GCCAAATTCACACGCCATTTGGATATGATACCTTTAAAGACACTTAACCTTATATCAGC						
GCCAAATTCACACGCCATTTGGATATGATACCTTTAAAGACACTTAACCTTATATCAGC						
1940	1950	1960	1970	1980	1990	2000
GCCAAATTCACACGCCATTTGGATATGATACCTTTAAAGACACTTAACCTTATATCAGC						
GCCAAATTCACACGCCATTTGGATATGATACCTTTAAAGACACTTAACCTTATATCAGC						

2. US-10-055-624B-1 (1-1857)

US-10-055-624B-3 Sequence 3, Application US/10055624B

Initial Score = 1633 Optimized Score = 1692 Significance = 2.53

Residue Identity = 93% Matches = 1718 Mismatches = 96

Gaps = 24 Conservative Substitutions = 0

X	10	20	30	40	50	60
GTGAATTTGCCCTT-----TCAAATGATCCGGAATGTGCGAAATACCTCGACGCGGCTTGGCTCTGTGT						
GTGAATTTGCCCTTGGCATAAATGATCCGGAATGTGCGAAATACCTCGACGCGGCTTGGCTCTGTGT						
X	10	20	30	40	50	60
CTCCAAACCGGACATCCCGGAAGTCAGCTTACGTCGCGGGAACACACAGTGCATGCGAGATCGGCGAGGAT						
CTCCAAACCGGACATCCCGGAAGTCAGCTTACGTCGCGGGAACACACAGTGCATGCGAGATCGGCGAGGAT						
80	90	100	110	120	130	140
CGGAAACCCACGAGGTGAGATGAGCCGCTGCTGTCTCCCTCCACACCAACCAACGCTGTGAGCTGAG						
CGGAAACCCACGAGGTGAGATGAGCCGCTGCTGTCTCCCTCCACACCAACCAACGCTGTGAGCTGAG						
150	160	170	180	190	200	210
CGTGGGAGCTGCTCAAGTGAATGACACACGAGCTTGTCTGTGCGGAGAGGACACGCTGTGCTGCCCTG						
CGTGGGAGCTGCTCAAGTGAATGACACACGAGCTTGTCTGTGCGGAGAGGACACGCTGTGCTGCCCTG						
220	230	240	250	260	270	280
TCGCAACGCTTCATGATGACATCTATTTTGACACCCATTAGTGTGGAACAGTGTGAATATCAAGGC						
TCGCAACGCTTCATGATGACATCTATTTTGACACCCATTAGTGTGGAACAGTGTGAATATCAAGGC						
290	300	310	320	330	340	350
CAAGTGAACCGGCGCTTCACTCCAGCATGAGTGGGAGCTCAGGTGGCTCGGAGGACCTGTGCTGA						
CAAGTGAACCGGCGCTTCACTCCAGCATGAGTGGGAGCTCAGGTGGCTCGGAGGACCTGTGCTGA						
360	370	380	390	400	410	420
CAAGTGAACCGGCGCTTCACTCCAGCATGAGTGGGAGCTCAGGTGGCTCGGAGGACCTGTGCTGA						
CAAGTGAACCGGCGCTTCACTCCAGCATGAGTGGGAGCTCAGGTGGCTCGGAGGACCTGTGCTGA						
370	380	390	400	410	420	430
CAAGTGAACCGGCGCTTCACTCCAGCATGAGTGGGAGCTCAGGTGGCTCGGAGGACCTGTGCTGA						
CAAGTGAACCGGCGCTTCACTCCAGCATGAGTGGGAGCTCAGGTGGCTCGGAGGACCTGTGCTGA						

[illegible]

[illegible]

Tue Aug 24 10:21:55 2004

us-10-055-624b-1.res

Page 5

1950 1960 1970 1980

Blank Sheet

580	590	600	610	620	630	640
TGTTATGACAGACCATCAAGACCTCTGGCCAACTGGCCATTGAGGCGATCTGGAGAGAGACT						
TGTTATGACAGACCATCAAGACCTCTGGCCAACTGGCCATTGAGGCGATCTGGAGAGAGACT						
TGTTATGACAGACCATCAAGACCTCTGGCCAACTGGCCATTGAGGCGATCTGGAGAGAGACT						
580	590	600	610	620	630	640

[illegible][illegible]

160 AGTGCAGATGAGCAGAGTGTGCTGCTCCCTGACACACACAAAGTGTGTGAGAGTGTGGGGAGCTGC
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
460
470
480
490
500
510
520
530
540
550
560
570
580
590
600
610
620
630
640
650
660
670
680
690
700
710
720
730
740
750
760
770
780
790
800
810
820
830
840
850
860
870
880
890
900
910
920
930
940
950
960
970
980
990
1000
1010
1020
1030
1040
1050
1060
1070
1080
1090
1100
1110
1120
1130
1140
1150
1160
1170
1180
1190
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
1300
1310
1320
1330
1340
1350
1360
1370
1380
1390
1400
1410
1420
1430
1440
1450
1460
1470
1480
1490
1500
1510
1520
1530
1540
1550
1560
1570
1580
1590
1600
1610
1620
1630
1640
1650
1660
1670
1680
1690
1700
1710
1720
1730
1740
1750
1760
1770
1780
1790
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
2040
2050
2060
2070
2080
2090
2100
2110
2120
2130
2140
2150
2160
2170
2180
2190
2200
2210
2220
2230
2240
2250
2260
2270
2280
2290
2300
2310
2320
2330
2340
2350
2360
2370
2380
2390
2400
2410
2420
2430
2440
2450
2460
2470
2480
2490
2500
2510
2520
2530
2540
2550
2560
2570
2580
2590
2600
2610
2620
2630
2640
2650
2660
2670
2680
2690
2700
2710
2720
2730
2740
2750
2760
2770
2780
2790
2800
2810
2820
2830
2840
2850
2860
2870
2880
2890
2900
2910
2920
2930
2940
2950
2960
2970
2980
2990
3000
3010
3020
3030
3040
3050
3060
3070
3080
3090
3100
3110
3120
3130
3140
3150
3160
3170
3180
3190
3200
3210
3220
3230
3240
3250
3260
3270
3280
3290
3300
3310
3320
3330
3340
3350
3360
3370
3380
3390
3400
3410
3420
3430
3440
3450
3460
3470
3480
3490
3500
3510
3520
3530
3540
3550
3560
3570
3580
3590
3600
3610
3620
3630
3640
3650
3660
3670
3680
3690
3700
3710
3720
3730
3740
3750
3760
3770
3780
3790
3800
3810
3820
3830
3840
3850
3860
3870
3880
3890
3900
3910
3920
3930
3940
3950
3960
3970
3980
3990
4000
4010
4020
4030
4040
4050
4060
4070
4080
4090
4100
4110
4120
4130
4140
4150
4160
4170
4180
4190
4200
4210
4220
4230
4240
4250
4260
4270
4280
4290
4300
4310
4320
4330
4340
4350
4360
4370
4380
4390
4400
4410
4420
4430
4440
4450
4460
4470
4480
4490
4500
4510
4520
4530
4540
4550
4560
4570
4580
4590
4600
4610
4620
4630
4640
4650
4660
4670
4680
4690
4700
4710
4720
4730
4740
4750
4760
4770
4780
4790
4800
4810
4820
4830
4840
4850
4860
4870
4880
4890
4900
4910
4920
4930
4940
4950
4960
4970
4980
4990
5000
5010
5020
5030
5040
5050
5060
5070
5080
5090
5100
5110
5120
5130
5140
5150
5160
5170
5180
5190
5200
5210
5220
5230
5240
5250
5260
5270
5280
5290
5300
5310
5320
5330
5340
5350
5360
5370
5380
5390
5400
5410
5420
5430
5440
5450
5460
5470
5480
5490
5500
5510
5520
5530
5540
5550
5560
5570
5580
5590
5600
5610
5620
5630
5640
5650
5660
5670
5680
5690
5700
5710
5720
5730
5740
5750
5760
5770
5780
5790
5800
5810
5820
5830
5840
5850
5860
5870
5880
5890
5900
5910
5920
5930
5940
5950
5960
5970
5980
5990
6000
6010
6020
6030
6040
6050
6060
6070
6080
6090
6100
6110
6120
6130
6140
6150
6160
6170
6180
6190
6200
6210
6220
6230
6240
6250
6260
6270
6280
6290
6300
6310
6320
6330
6340
6350
6360
6370
6380
6390
6400
6410
6420
6430
6440
6450
6460
6470
6480
6490
6500
6510
6520
6530
6540
6550
6560
6570
6580
6590
6600
6610
6620
6630
6640
6650
6660
6670
6680
6690
6700
6710
6720
6730
6740
6750
6760
6770
6780
6790
6800
6810
6820
6830
6840
6850
6860
6870
6880
6890
6900
6910
6920
6930
6940
6950
6960
6970
6980
6990
7000
7010
7020
7030
7040
7050
7060
7070
7080
7090
7100
7110
7120
7130
7140
7150
7160
7170
7180
7190
7200
7210
7220
7230
7240
7250
7260
7270
7280
7290
7300
7310
7320
7330
7340
7350
7360
7370
7380
7390
7400
7410
7420
7430
7440
7450
7460
7470
7480
7490
7500
7510
7520
7530
7540
7550
7560
7570
7580
7590
7600
7610
7620
7630
7640
7650
7660
7670
7680
7690
7700
7710
7720
7730
7740
7750
7760
7770
7780
7790
7800
7810
7820
7830
7840
7850
7860
7870
7880
7890
7900
7910
7920
7930
7940
7950
7960
7970
7980
7990
8000
8010
8020
8030
8040
8050
8060
8070
8080
8090
8100
8110
8120
8130
8140
8150
8160
8170
8180
8190
8200
8210
8220
8230
8240
8250
8260
8270
8280
8290
8300
8310
8320
8330
8340
8350
8360
8370
8380
8390
8400
8410
8420
8430
8440
8450
8460
8470
8480
8490
8500
8510
8520
8530
8540
8550
8560
8570
8580
8590
8600
8610
8620
8630
8640
8650
8660
8670
8680
8690
8700
8710
8720
8730
8740
8750
8760
8770
8780
8790
8800
8810
8820
8830
8840
8850
8860
8870
8880
8890
8900
8910
8920
8930
8940
8950
8960
8970
8980
8990
9000
9010
9020
9030
9040
9050
9060
9070
9080
9090
9100
9110
9120
9130
9140
9150
9160
9170
9180
9190
9200
9210
9220
9230
9240
9250
9260
9270
9280
9290
9300
9310
9320
9330
9340
9350
9360
9370
9380
9390
9400
9410
9420
9430
9440
9450
9460
9470
9480
9490
9500
9510
9520
9530
9540
9550
9560
9570
9580
9590
9600
9610
9620
9630
9640
9650
9660
9670
9680
9690
9700
9710
9720
9730
9740
9750
9760
9770
9780
9790
9800
9810
9820
9830
9840
9850
9860
9870
9880
9890
9900
9910
9920
9930
9940
9950
9960
9970
9980
9990
10000
10010
10020
10030
10040
10050
10060
10070
10080
10090
10100
10110
10120
10130
10140
10150
10160
10170
10180
10190
10200
10210
10220
10230
10240
10250
10260
10270
10280
10290
10300
10310
10320
10330
10340
10350
10360
10370
10380
10390
10400
10410
10420
10430
10440
10450
10460
10470
10480
10490
10500
10510
10520
10530
10540
10550
10560
10570
10580
10590
10600
10610
10620
10630
10640
10650
10660
10670
10680
10690
10700
10710
10720
10730
10740
10750
10760
10770
10780
10790
10800
10810
10820
10830
10840
10850
10860
10870
10880
10890
10900
10910
10920
10930
10940
10950
10960
10970
10980
10990
11000
11010
11020
11030
11040
11050
11060
11070
11080
11090
11100
11110
11120
11130
11140
11150
11160
11170
11180
11190
11200
11210
11220
11230
11240
11250
11260
11270
11280
11290
11300
11310
11320
11330
11340
11350
11360
11370
11380
11390
11400
11410
11420
11430
11440
11450
11460
11470
11480
11490
11500
11510
11520
11530
11540
11550
11560
11570
11580
11590
11600
11610
11620
11630
11640
11650
11660
11670
11680
11690
11700
11710
11720
11730
11740
11750
11760
11770
11780
11790
11800
11810
11820
11830
11840
11850
11860
11870
11880
11890
11900
11910
11920
11930
11940
11950
11960
11970
11980
11990
12000
12010
12020
12030
12040
12050
12060
12070
12080
12090
12100
12110
12120
12130
12140
12150
12160
12170
12180
12190
12200
12210
12220
12230
12240
12250
12260
12270
12280
12290
12300
12310
12320
12330
12340
12350
12360
12370
12380
12390
12400
12410
12420
12430
12440
12450
12460
12470
12480
12490
12500
12510
12520
12530
12540
12550
12560
12570
12580
12590
12600
12610
12620
12630
12640
12650
12660
12670
12680
12690
12700
12710
12720
12730
12740
12750
12760
12770
12780
12790
12800
12810
12820
12830
12840
12850
12860
12870
12880
12890
12900
12910
12920
12930
12940
12950
12960
12970
12980
12990
13000
13010
13020
13030
13040
13050
13060
13070
13080
13090
13100
13110
13120
13130
13140
13150
13160
13170
13180
13190
13200
13210
13220
13230
13240
13250
13260
13270
13280
13290
13300
13310
13320
13330
13340
13350
13360
13370
13380
13390
13400
13410
13420
13430
13440
13450
13460
13470
13480
13490
13500
13510
13520
13530
13540
13550
13560
13570
13580
13590
13600
13610
13620
13630
13640
13650
13660
13670
13680
13690
13700
13710
13720
13730
13740
13750
13760
13770
13780
13790
13800
13810
13820
13830
13840
13850
13860
13870
13880
13890
13900
13910
13920
13930
13940
13950
13960
13970
13980
13990
14000
14010
14020
14030
14040
14050
14060
14070
14080
14090
14100
14110
14120
14130
14140
14150
14160
14170
14180
14190
14200
14210
14220
14230
14240
14250
14260
14270
14280
14290
14300
14310
14320
14330
14340
14350
14360
14370
14380
14390
14400
14410
14420
14430
14440
14450
14460
14470
14480
14490
14500
14510
14520
14530
14540
14550
14560
14570
14580
14590
14600
14610
14620
14630
14640
14650
14660
14670
14680
14690
14700
14710
14720
14730
14740
14750
14760
14770
14780
14790
14800
14810
14820
14830
14840
14850
14860
14870
14880
14890
14900
14910
14920
14930
14940
14950
14960
14970
14980
14990
15000
15010
15020
15030
15040
15050
15060
15070
15080
15090
15100
15110
15120
15130
15140
15150
15160
15170
15180
15190
15200
15210
15220
15230
15240
15250
15260
15270
15280
15290
15300
15310
15320
15330
15340
15350
15360
15370
15380
15390
15400
15410
15420
15430
15440
15450
15460
15470
15480
15490
15500
15510
15520
15530
15540
15550
15560
15570
15580
15590
15600
15610
15620
15630
15640
15650
15660
15670
15680
15690
15700
15710
15720
15730
15740
15750
15760
15770
15780
15790
15800
15810
15820
15830
15840
15850
15860
15870
15880
15890
15900
15910
15920
15930
15940
15950
15960
15970
15980
15990
16000
16010
16020
16030
16040
16050
16060
16070
16080
16090
16100
16110
16120
16130
16140
16150
16160
16170
16180
16190
16200
16210
16220
16230
16240
16250
16260
16270
16280
16290
16300
16310
16320
16330
16340
16350
16360
16370
16380
16390
16400
16410
16420
16430
16440
16450
16460
16470
16480
16490
16500
16510
16520
16530
16540
16550
16560
16570
16580
16590
16600
16610
16620
16630
16640
16650
16660
16670
16680
16690
16700
16710
16720
16730
16740
16750
16760
16770
16780
16790
16800
16810
16820
16830
16840
16850
16860
16870
16880
16890
16900
16910
16920
16930
16940
16950
16960
16970
16980
16990
17000
17010
17020
17030
17040
17050
17060
17070
17080
17090
17100
17110
17120
17130
17140
17150
17160
17170
17180
17190
17200
17210
17220
17230
17240
17250
17260
17270
17280
17290
17300
17310
17320
17330
17340
17350
17360
17370
17380
17390
17400
17410
17420
17430
17440
17450
17460
17470
17480
17490
17500
17510
17520
17530
17540
17550
17560
17570
17580
17590
17600
17610
17620
17630
17640
17650
17660
17670
17680
17690
17700
17710
17720
17730
17740
17750
17760
17770
17780
17790
17800
17810
17820
17830
17840
17850
17860
17870
17880
17890
17900
17910
17920
17930
17940
17950
17960
17970
17980
17990
18000
18010
18020
18030
18040
18050
18060
18070
18080
18090
18100
18110
18120
18130
18140
18150
18160
18170
18180
18190
18200
18210
18220
18230
18240
18250
18260
18270
18280
18290
18300
18310
18320
18330
18340
18350
18360
18370
18380
18390
18400
18410
18420
18430
18440
18450
18460
18470
18480
18490
18500
18510
18520
18530
18540
18550
18560
18570
18580
18590
18600
18610
18620
18630
18640
18650
18660
18670
18680
18690
18700
18710
18720
18730
18740
18750
18760
18770
18780
18790
18800
18810
18820
18830
18840
18850
18860
18870
18880
18890
18900
18910
18920
18930
18940
18950
18960
18970
18980
18990
19000
19010
19020
19030
19040
19050
19060
19070
19080
19090

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: August 12, 2004, 15:45:27 ; Search time 49 Seconds
(without alignments)
3888.853 Million cell updates/sec

Title: US-10-055-624b-2

Perfect score: 3181
Sequence: 1 MIONVGNHRLRGSLAVFSNR.....TLKAPVAKINSRFGYLDPT 607

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3181	100.0	607	US-10-055-624b-2	Sequence 2, Appl1
2	2838	89.2	594	US-10-055-624b-4	Sequence 4, Appl1
3	2530	79.5	594	US-10-055-624b-6	Sequence 6, Appl1
4	1242.5	39.1	555	US-10-336-472-50	Sequence 59, Appl1
5	413	13.0	119	US-09-925-298-599	Sequence 599, App
6	413	13.0	119	US-10-102-806-599	Sequence 599, App
7	255	8.0	339	US-09-738-626-644	Sequence 634, App
8	255	8.0	339	US-10-627-476-230	Sequence 230, App
9	223.5	7.0	155	US-10-282-122A-54710	Sequence 54710, A
10	223.5	7.0	155	US-10-289-762-493	Sequence 693, App
11	220	6.9	339	US-09-809-391-488	Sequence 488, App
12	220	6.9	339	US-09-882-171-488	Sequence 488, App
13	220	6.9	339	US-10-164-861-488	Sequence 488, App
14	219	6.9	370	US-10-094-749-2967	Sequence 2967, App
15	207.5	6.5	159	US-10-282-122A-55313	Sequence 55313, A

15	203	6.4	168	12	US-10-282-122A-45753	Sequence 45753, A
16	201	6.3	207	14	US-10-156-761-12563	Sequence 12563, A
17	186	5.8	156	12	US-10-282-122A-45621	Sequence 45621, A
18	185	5.7	176	12	US-10-282-122A-70623	Sequence 70623, A
19	182	5.7	455	10	US-09-906-408A-8	Sequence 8, Appl1
20	181.5	5.6	183	12	US-10-282-122A-63369	Sequence 63369, A
21	177	5.6	174	12	US-10-282-122A-58829	Sequence 58829, A
22	177	5.6	174	12	US-10-335-977-8306	Sequence 8306, App
23	177	5.6	177	12	US-10-335-977-8307	Sequence 8307, App
24	177	5.6	172	12	US-10-282-122A-60467	Sequence 60467, A
25	173	5.4	170	12	US-10-282-122A-45497	Sequence 45497, A
26	171.5	5.4	177	12	US-10-282-122A-71369	Sequence 71369, A
27	171.5	5.4	177	12	US-10-437-963-119373	Sequence 119373, A
28	171.5	5.4	526	16	US-10-282-122A-70386	Sequence 70386, A
29	170.5	5.2	176	12	US-10-282-122A-56911	Sequence 56911, A
30	166	5.0	178	12	US-10-282-122A-56911	Sequence 56911, A
31	159.5	4.6	434	12	US-10-424-599-151287	Sequence 151287, A
32	146.5	4.6	321	12	US-10-092-900A-222	Sequence 222, App
33	146.5	4.4	110	12	US-10-424-599-220213	Sequence 220213, A
34	140.5	4.4	438	10	US-09-906-408A-7	Sequence 7, Appl1
35	140.5	4.4	137	12	US-10-282-122A-50648	Sequence 50648, A
36	139.5	4.4	131	12	US-10-282-122A-49795	Sequence 49795, A
37	133	4.2	137	12	US-10-282-122A-54459	Sequence 54459, A
38	131	4.0	136	16	US-10-408-765A-2655	Sequence 2655, A
39	128	4.0	136	12	US-10-282-122A-51198	Sequence 51198, A
40	124.5	3.9	136	12	US-10-282-122A-48235	Sequence 48235, A
41	121	3.8	2108	8	US-08-973-363-16	Sequence 16, Appl1
42	119.5	3.8	140	12	US-10-282-122A-44586	Sequence 44586, A
43	118	3.7	445	14	US-10-177-293-296	Sequence 296, App
44	118	3.7	445	15	US-10-435-696-33	Sequence 33, Appl1
45	115	3.6	1289	9	US-09-932-145-11	Sequence 11, Appl1

ALIGNMENTS

RESULT 1
US-10-055-624b-2
; Sequence 2, Application US/10055624B
; Publication No. US20030220238A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Sean H
; APPLICANT: Chui, Charissa
; APPLICANT: Goddard, Audrey D
; TITLE OF INVENTION: BFTT COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 9800081-0066
; CURRENT APPLICATION NUMBER: US/10/055,624B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/263,362
; PRIOR FILING DATE: 2002-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-055-624b-2

Query Match	Similarity	Score	Pred. No.	DB	Length	Indels	Gaps
Best Local	100.0%	3181	0	15	607	0	0
Matches	607	Conservative	0	Mismatches	0	Indels	0
DB	1	MIONVGNHRLRGSLAVFSNR	SRKSAIRAGDSAMADEGYRNPTEVQMSQVLJECHTNQ	60			
QY	1	MIONVGNHRLRGSLAVFSNR	SRKSAIRAGDSAMADEGYRNPTEVQMSQVLJECHTNQ	60			
DB	61	RGEISVGGLKMDITTA	CLSAERHAGCPVTASMDIYFEHTISVGQVNNIKAKNRAFN	120			
QY	61	RGEISVGGLKMDITTA	CLSAERHAGCPVTASMDIYFEHTISVGQVNNIKAKNRAFN	120			
DB	61	RGEISVGGLKMDITTA	CLSAERHAGCPVTASMDIYFEHTISVGQVNNIKAKNRAFN	120			
QY	121	SSMEVGIOVASEDDLCSEKQNVCKALATFVARREITTKKQITRTEEBKMEHSVAAR	180				
DB	121	SSMEVGIOVASEDDLCSEKQNVCKALATFVARREITTKKQITRTEEBKMEHSVAAR	180				
QY	121	SSMEVGIOVASEDDLCSEKQNVCKALATFVARREITTKKQITRTEEBKMEHSVAAR	180				
DB	121	SSMEVGIOVASEDDLCSEKQNVCKALATFVARREITTKKQITRTEEBKMEHSVAAR	180				

```

QY 181 RRMRLVYADTTIKDLANCAIQGDLBSRDCSRWVPAKTRVESVELVLPHPANHOGNTFGG 240
Db 181 RRMRLVYADTTIKDLANCAIQGDLBSRDCSRWVPAKTRVESVELVLPHPANHOGNTFGG 240
QY 241 QIMAMMENAVATTIASAASLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMVEGV 300
Db 241 QIMAMMENAVATTIASAASLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMVEGV 300
QY 301 CVEAYROEAEFTHRRHINSAFMTFVVLADDDQPOLPMIRPOGDDGRRRREASARKKIRL 360
Db 301 CVEAYROEAEFTHRRHINSAFMTFVVLADDDQPOLPMIRPOGDDGRRRREASARKKIRL 360
QY 361 DRKYIVSCKQTEVPLSVPMDSNQYVLSYNNVSSIKMLVAKDNWVLSSEISQVRLYTTLED 420
Db 361 DRKYIVSCKQTEVPLSVPMDSNQYVLSYNNVSSIKMLVAKDNWVLSSEISQVRLYTTLED 420
QY 421 DKFLSFHMEMVHVDAQAQAFLLISDLRORPEMDKHYRSVELVQVDEDDAIYHVTSPALG 480
Db 421 DKFLSFHMEMVHVDAQAQAFLLISDLRORPEMDKHYRSVELVQVDEDDAIYHVTSPALG 480
QY 481 GHTKPODFVLIASRRKPCDNDPVYIALRSVTLPTHRETPERYRGETLCSGFCIMRBGDQ 540
Db 481 GHTKPODFVLIASRRKPCDNDPVYIALRSVTLPTHRETPERYRGETLCSGFCIMRBGDQ 540
QY 541 LTKCCKWRVSLTEIVSASGFYSNGLBSRSGKRSRSDGNGKLAGHLSITLKAIPVAKINSR 600
Db 541 LTKCCKWRVSLTEIVSASGFYSNGLBSRSGKRSRSDGNGKLAGHLSITLKAIPVAKINSR 600
QY 601 FGVIQDT 607
Db 601 FGVIQDT 607

```

RESULT 2

```

US-10-055-624B-4
; Sequence 4, Application US/10055624B
; Publication No. US2003022038A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Sean H
; APPLICANT: Chui, Clarissa
; APPLICANT: Goddard, Audrey D
; TITLE OF INVENTION: BIT COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 9800081-0066
; CURRENT APPLICATION NUMBER: US/10/055,624B
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/263,362
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-055-624B-4

```

Query Match 89.2%; Score 2838; DB 15; Length 594;
 Best Local Similarity 100.0%; Pred. No. 2e-283;
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MIQVGNHILRGGLASVFSNRTSRKSAIRAGNDSAMADGEGYRNPTVOMSOVLPCHTNQ 60
Db 1 MIQVGNHILRGGLASVFSNRTSRKSAIRAGNDSAMADGEGYRNPTVOMSOVLPCHTNQ 60
QY 61 RGEISVQQLIKMIDTTACLSAERHAGCPCVTASMDIYFEHTISVGVNNIKAVNAFN 120
Db 61 RGEISVQQLIKMIDTTACLSAERHAGCPCVTASMDIYFEHTISVGVNNIKAVNAFN 120
QY 121 SSMVEGIVQVASEDLCSEKQNNVCKALATFVARREITVKLKOITPREEKMEHSAER 180
Db 121 SSMVEGIVQVASEDLCSEKQNNVCKALATFVARREITVKLKOITPREEKMEHSAER 180
QY 180 SSMVEGIVQVASEDLCSEKQNNVCKALATFVARREITVKLKOITPREEKMEHSAER 180

```

```

QY 181 RRMRLVYADTTIKDLANCAIQGDLBSRDCSRWVPAKTRVESVELVLPHPANHOGNTFGG 240
Db 181 RRMRLVYADTTIKDLANCAIQGDLBSRDCSRWVPAKTRVESVELVLPHPANHOGNTFGG 240
QY 241 QIMAMMENAVATTIASAASLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMVEGV 300
Db 241 QIMAMMENAVATTIASAASLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMVEGV 300
QY 301 CVEAYROEAEFTHRRHINSAFMTFVVLADDDQPOLPMIRPOGDDGRRRREASARKKIRL 360
Db 301 CVEAYROEAEFTHRRHINSAFMTFVVLADDDQPOLPMIRPOGDDGRRRREASARKKIRL 360
QY 361 DRKYIVSCKQTEVPLSVPMDSNQYVLSYNNVSSIKMLVAKDNWVLSSEISQVRLYTTLED 420
Db 361 DRKYIVSCKQTEVPLSVPMDSNQYVLSYNNVSSIKMLVAKDNWVLSSEISQVRLYTTLED 420
QY 421 DKFLSFHMEMVHVDAQAQAFLLISDLRORPEMDKHYRSVELVQVDEDDAIYHVTSPALG 480
Db 421 DKFLSFHMEMVHVDAQAQAFLLISDLRORPEMDKHYRSVELVQVDEDDAIYHVTSPALG 480
QY 481 GHTKPODFVLIASRRKPCDNDPVYIALRSVTLPTHRETPERYRGETLCSGFCIMRBGDQ 540
Db 481 GHTKPODFVLIASRRKPCDNDPVYIALRSVTLPTHRETPERYRGETLCSGFCIMRBGDQ 540
QY 541 LTK 543
Db 541 LTK 543

```

RESULT 3

```

US-10-055-624B-6
; Sequence 6, Application US/10055624B
; Publication No. US2003022038A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Sean H
; APPLICANT: Chui, Clarissa
; APPLICANT: Goddard, Audrey D
; TITLE OF INVENTION: BIT COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 9800081-0066
; CURRENT APPLICATION NUMBER: US/10/055,624B
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/263,362
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-055-624B-6

```

Query Match 79.5%; Score 2530; DB 15; Length 594;
 Best Local Similarity 81.9%; Pred. No. 1.3e-251;
 Matches 493; Conservative 35; Mismatches 54; Indels 20; Gaps 6;

```

QY 1 MIQVGNHILRGGLASVFSNRTSRKSAIRAGNDSAMADGEGYRNPTVOMSOVLPCHTNQ 58
Db 1 MIQVGNHILRGGLASVFSNRTSRKSAIRAGNDSAMADGEGYRNPTVOMSOVLPCHTNQ 58
QY 59 NQRELSIGQLIKMIDTTACLSAERHAGCPCVTASMDIYFEHTISVGVNNIKAVNAFN 118
Db 59 NQRELSIGQLIKMIDTTACLSAERHAGCPCVTASMDIYFEHTISVGVNNIKAVNAFN 118
QY 119 FNSMEVGIVQVASEDLCSEKQNNVCKALATFVARREITVKLKOITPREEKMEHSAER 178
Db 119 FNSMEVGIVQVASEDLCSEKQNNVCKALATFVARREITVKLKOITPREEKMEHSAER 178
QY 179 ERRRMLVYADTTIKDLANCAIQGDLBSRDCSRWVPAKTRVESVELVLPHPANHOGNTF 238
Db 179 ERRRMLVYADTTIKDLANCAIQGDLBSRDCSRWVPAKTRVESVELVLPHPANHOGNTF 238
QY 239 GQIMAMMENAVATTIASAASLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMVEV 298
Db 239 GQIMAMMENAVATTIASAASLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMVEV 298

```

```
Db 240 GGGIMAMMENVAATIAASRLCHAHPTLKALIMFHFRRPSQVGDVLKAVINNAFKHSMEX 299
Qy 299 GVCVEAYROEATTHRRHINSAPMTFVVLADADDPQLLPWIRPQGGERRYRASARKKI 358
Db 300 GVCVEAYROEATTHRRHINSAPMTFVVLADADDPQLLPWIRPQGGERRYRASARKKI 359
Qy 359 RLDKRYIVSCQTEVPLSVPMPSNOVYLSYNNVSSIKMLVAKDNVLSSELSQVRLYTL 418
Db 360 RLDKRYIVSCQTEVPLSVPMPSNOVYLSYNNVSSIKMLVAKDNVLSSELSQVRLYTL 419
Qy 419 EDDKESFHEMRYVYVDAQAFLILSDLRORPEWDXHYSVEIYQVDEDDAIYHVTSPA 478
Db 420 EED-FLSPHEMRYVYVDAQAFLILSDLRORPEWDXHYSVEIYQVDEDDAIYHVTSPA 478
Qy 479 LGGHTKQDPFVILASRRKPCDNDPVIYALRSVTLPTHRETPEYRRGETLCSGFCIMRBS 538
Db 479 LGGHTKQDPFVILASRRKPCDNDPVIYALRSVTLPTHRETPEYRRGETLCSGFCIMRBS 538
Qy 539 DDLTKCQWVRVSLTELVSASGFYSWGLSRSKGRSDGNNGKLAGHLSLTKAIPYAKIN 598
Db 539 DDLTKCQWVRVSLTELVSASGFYSWGLSRSKGRSDGNNGKLAGHLSLTKAIPYAKIN 598
Qy 599 SR 600
Db 599 SR 600
Qy 583 NR 584
Db 583 NR 584

RESULT 4
US-10-336-472-50
; Sequence 50, Application US/10336472
; Publication No. US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangoli, Rasha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Miller, Isabelle
; APPLICANT: Padigaru, Moraliadhara
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Spaderina, Steven K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Orc, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zernhusen, Bryan D.
; APPLICANT: Zhong, Mei
```

```
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuroSeqList version 0.1
; SEQ ID NO 50
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-50

Query Match 39.1%; Score 1242.5; DB 12; Length 555;
Best Local Similarity 52.2%; Pred. No. 1.2e-118;
Matches 260; Conservative 82; Mismatches 143; Indels 13; Gaps 7;

Qy 44 PTEVQMSQVLPCHTNORGELSVGQLKWIIDTACLSARHAGCPCTASMDIYENTL 103
Db 6 PGEVWMSQAIORPHANARBELSAGQLKWIIDTACLSARHAGCPCTASMDIYENTL 65
Qy 104 SVGQVNVIRAKVNRAPNSSMEVGIQVASEDLCEKQMNVCALATVAVR--REITRYKL 160
Db 66 RVGQVITIKAKVTRAFSTMEISIKVMQDMLTGIEKIVSVAFSTVAKVGRKE--KIHL 123
Qy 161 KQITPREBKKEHSAAREBRMRVYADPTIDOLANCAIQGLSRRDSCRMVPAKTRY 220
Db 124 KPVTLITLBDVHEHNIAAERRKVRLOHEDTFNNLMKSSKFDLLIFDEEGAVSTGTSV 183
Qy 221 ESVEIVLPFHANHGNTFGQIMAMENVAITIAASRLCHAHPTLKALIMFHFRRPSQVGD 280
Db 184 QGIEIVLPFHANHGNTFGQIMAMENVAITIAASRLCHAHPTLKALIMFHFRRPSQVGD 243
Qy 281 RLVKATVNNAFKHSNEVGVCEAY--ROEATTHRRHINSAPMTFVVLADADDPQL--P 336
Db 244 RLVEFTAVNNTFQTCVEGVVEAFDQCEMARGRHINSAPF--TNSADDEKENLITPP 301
Qy 337 WIRPQGGERRYRRASARKKIRLDKRYIVSCQTEVPLSVPMPSNOVYLSYNNVSLK 396
Db 302 RIQPLSKDPPRRKRGAIARKRIRLGRKVIYSHR--EVLICIMWDISKQSLSSNVAALK 360
Qy 397 MLVAKDNVLSSELSQVRLYTLDEDKELSFHEMRYVYVDAQAFLILSDLRORPEWDXH 456
Db 361 KLAARKGMEVTVSEKIKITYLEHDVLSVMVEKIVGSAHLAYRLSDPTKRLMDPHF 420
Qy 457 KSVELVQVDEDDAIYHVTSPAIGHTKQDPFVILASRRKPCDNDPVIYALRSVTLPTH 516
Db 421 VSCVEIVDWSBDQHLHITCPL--NDDKPKDLVILVSRKPKLDGNTYTVAVXSVLLPSV 479
Qy 517 RETPEYRRGETLCSGFCIL 534
Db 480 PPSQYITRSEITCAGFLI 497
```

	ORGANISM:	Homo sapiens
	FEATURE:	
	NAME/KEY:	SITE
	LOCATION:	(58)
	OTHER INFORMATION:	Xaa equals any of the naturally occurring L-amino acids
	NAME/KEY:	SITE
	LOCATION:	(68)
	OTHER INFORMATION:	Xaa equals any of the naturally occurring L-amino acids
	NAME/KEY:	SITE
	LOCATION:	(88)
	OTHER INFORMATION:	Xaa equals any of the naturally occurring L-amino acids
	NAME/KEY:	SITE
	LOCATION:	(98)
	OTHER INFORMATION:	Xaa equals any of the naturally occurring L-amino acids
	NAME/KEY:	SITE
	LOCATION:	(99)
	OTHER INFORMATION:	Xaa equals any of the naturally occurring L-amino acids
	US-10-102-806-599	
QY	Query Match	13.0%; Score 413; DB 14; Length 119;
	Best Local Similarity	96.4%; Pred. No. 4.6e-34;
D8	Matches	81; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY	383	NOYLIVNNVNSLSIKMVIYAKNWNVLSSRIISQVRLTYTLEDDKFLSFMHMYVHVDAQAQEL 442
D8	4	DOYLSLNNVNSLSLMVLAKDNWVLSSISQVRLTYTLEDDKFLSFMHMYVHVDAQAQFLL 63
QY	443	LSDLRORPEWDKHYSVELVQOVD 466
D8	64	LSDLXGRPEWDKHYSVELVQOVD 87
	RESULT 7	
	US-09-738-626-6344	
	Sequence 6344, Application US/09738626	
	Publication No. US20020197605A1	
	GENERAL INFORMATION:	
	APPLICANT: NAYAGAWA, SATOSHI	
	APPLICANT: MIZOGUCHI, HIROSHI	
	APPLICANT: ANDO, SEIKO	
	APPLICANT: HAYASHI, MIKIRO	
	APPLICANT: OCHIAI, KEIKO	
	APPLICANT: YOKOI, HARUHIKO	
	APPLICANT: TATEISHI, NAOKO	
	APPLICANT: SENOH, AKIHITO	
	APPLICANT: IKEDA, MASATO	
	APPLICANT: OZAKI, AKIO	
	TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES	
	FILE REFERENCE: 249-125	
	CURRENT APPLICATION NUMBER: US/09/738, 626	
	PRIOR FILING DATE: 2000-12-18	
	PRIOR APPLICATION NUMBER: JP 99/377484	
	PRIOR FILING DATE: 1999-12-16	
	PRIOR APPLICATION NUMBER: JP 00/159162	
	PRIOR FILING DATE: 2000-04-07	
	PRIOR APPLICATION NUMBER: JP 00/280988	
	PRIOR FILING DATE: 2000-08-03	
	NUMBER OF SEQ ID NOS: 7059	
	SOFTWARE: PatentIn ver. 3.0	
	SEQ ID NO 6344	
	LENGTH: 339	
	TYPE: PRT	
	ORGANISM: Corynebacterium glutamicum	
	US-09-738-626-6344	
QY	Query Match	8.0%; Score 255; DB 9; Length 339;
	Best Local Similarity	25.7%; Pred. No. 5.3e-17;
D8	Matches	89; Conservative 56; Mismatches 17; Indels 30; Gaps 10;
QY	41	YRNPTVEQMQLVLPCHTNORGELSY--GOLTKVIDTTACLSAEHRNAGCPCTASMDIY 98
D8	14	HRSP-EVTLRFMAAPTIVLAGSHGVGGKRVLEMIDKAAYACAIGMSGTTYCVTAIVGHIIH 72

RESULT 8
US-10-627-476-230
; Sequence 230, Application US/10627476
; Publication No. US20040030116A1
GENERAL INFORMATION

Query Match	8.0%;	Score 255;	DB 12;	Length 339;
Best Local Similarity	25.7%;	Pred. No. 5.3e-17;		
Matches	89;	Conservative	56;	Mismatches 171;
			Indels	30;
			Gaps	10.

QY 41 YRNPEVQMSQVLPCHTNQRGELSV--GQLKIIDITTCLSAERHAGCPVTASMDIY 98
 Db 14 HRSP-EVVLRFMAPTDVLMAGSHGVGGRLLEIDKAYACATQMSGTCVTATYGVGH 72

RESULT 9
US-10-282-122A-54710
; Sequence 54710, Application US/10282122A
; Publication No. US20040029129A1

RESULT 9
US-10-282-122A-54710
Sequence 54710, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Opsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA_034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
Number of rows

```

/ / SOFTWARE: PatentIn version 3.1
/ / SEQ ID NO: 54710
/ / LENGTH: 155
/ / TYPE: PRN
/ / ORGANISM: Chlamydia pneumoniae
US-10-283-122A-54710

Query Match      7.08; Score 223.5; DB 12; Length 155;
Best Local Similarity 35.58; Pred. NO. 2,7e-14;

```


Tue Aug 24 10:21:58 2004

us-10-055-624b-2.rapp

Page 7

1	PRIOR FILING DATE: 1997-05-23	1	PRIOR APPLICATION NUMBER: 60/047,615	1	PRIOR FILING DATE: 1997-08-22	1	PRIOR APPLICATION NUMBER: 60/056,889
2	PRIOR APPLICATION NUMBER: 60/047,615	2	PRIOR FILING DATE: 1997-05-23	2	PRIOR FILING DATE: 1997-08-22	2	PRIOR APPLICATION NUMBER: 60/056,893
3	PRIOR FILING DATE: 1997-05-23	3	PRIOR APPLICATION NUMBER: 60/047,597	3	PRIOR FILING DATE: 1997-08-22	3	PRIOR APPLICATION NUMBER: 60/056,630
4	PRIOR FILING DATE: 1997-05-23	4	PRIOR APPLICATION NUMBER: 60/047,502	4	PRIOR FILING DATE: 1997-08-22	4	PRIOR APPLICATION NUMBER: 60/056,878
5	PRIOR FILING DATE: 1997-05-23	5	PRIOR APPLICATION NUMBER: 60/047,633	5	PRIOR FILING DATE: 1997-08-22	5	PRIOR APPLICATION NUMBER: 60/056,662
6	PRIOR APPLICATION NUMBER: 60/047,633	6	PRIOR FILING DATE: 1997-05-23	6	PRIOR FILING DATE: 1997-08-22	6	PRIOR APPLICATION NUMBER: 60/056,872
7	PRIOR FILING DATE: 1997-05-23	7	PRIOR APPLICATION NUMBER: 60/047,583	7	PRIOR FILING DATE: 1997-08-22	7	PRIOR APPLICATION NUMBER: 60/056,882
8	PRIOR FILING DATE: 1997-05-23	8	PRIOR APPLICATION NUMBER: 60/047,617	8	PRIOR FILING DATE: 1997-08-22	8	PRIOR APPLICATION NUMBER: 60/056,911
9	PRIOR APPLICATION NUMBER: 60/047,617	9	PRIOR FILING DATE: 1997-05-23	9	PRIOR FILING DATE: 1997-08-22	9	PRIOR APPLICATION NUMBER: 60/056,874
10	PRIOR FILING DATE: 1997-05-23	10	PRIOR APPLICATION NUMBER: 60/047,592	10	PRIOR FILING DATE: 1997-08-22	10	PRIOR APPLICATION NUMBER: 60/056,888
11	PRIOR FILING DATE: 1997-05-23	11	PRIOR APPLICATION NUMBER: 60/047,581	11	PRIOR FILING DATE: 1997-08-22	11	PRIOR APPLICATION NUMBER: 60/056,910
12	PRIOR APPLICATION NUMBER: 60/047,581	12	PRIOR FILING DATE: 1997-05-23	12	PRIOR FILING DATE: 1997-08-22	12	PRIOR APPLICATION NUMBER: 60/056,864
13	PRIOR FILING DATE: 1997-05-23	13	PRIOR APPLICATION NUMBER: 60/047,584	13	PRIOR FILING DATE: 1997-08-22	13	PRIOR APPLICATION NUMBER: 60/056,874
14	PRIOR FILING DATE: 1997-05-23	14	PRIOR APPLICATION NUMBER: 60/047,500	14	PRIOR FILING DATE: 1997-08-22	14	PRIOR APPLICATION NUMBER: 60/056,894
15	PRIOR FILING DATE: 1997-05-23	15	PRIOR APPLICATION NUMBER: 60/047,587	15	PRIOR FILING DATE: 1997-08-22	15	PRIOR APPLICATION NUMBER: 60/056,911
16	PRIOR APPLICATION NUMBER: 60/047,587	16	PRIOR FILING DATE: 1997-05-23	16	PRIOR FILING DATE: 1997-08-22	16	PRIOR APPLICATION NUMBER: 60/056,874
17	PRIOR FILING DATE: 1997-05-23	17	PRIOR APPLICATION NUMBER: 60/047,492	17	PRIOR FILING DATE: 1997-08-22	17	PRIOR APPLICATION NUMBER: 60/056,874
18	PRIOR FILING DATE: 1997-05-23	18	PRIOR APPLICATION NUMBER: 60/047,598	18	PRIOR FILING DATE: 1997-08-22	18	PRIOR APPLICATION NUMBER: 60/056,874
19	PRIOR FILING DATE: 1997-05-23	19	PRIOR APPLICATION NUMBER: 60/047,613	19	PRIOR FILING DATE: 1997-08-22	19	PRIOR APPLICATION NUMBER: 60/056,910
20	PRIOR APPLICATION NUMBER: 60/047,613	20	PRIOR FILING DATE: 1997-05-23	20	PRIOR FILING DATE: 1997-08-22	20	PRIOR APPLICATION NUMBER: 60/056,864
21	PRIOR FILING DATE: 1997-05-23	21	PRIOR APPLICATION NUMBER: 60/047,582	21	PRIOR FILING DATE: 1997-08-22	21	PRIOR APPLICATION NUMBER: 60/056,874
22	PRIOR FILING DATE: 1997-05-23	22	PRIOR APPLICATION NUMBER: 60/047,596	22	PRIOR FILING DATE: 1997-08-22	22	PRIOR APPLICATION NUMBER: 60/056,874
23	PRIOR FILING DATE: 1997-05-23	23	PRIOR APPLICATION NUMBER: 60/047,612	23	PRIOR FILING DATE: 1997-08-22	23	PRIOR APPLICATION NUMBER: 60/056,874
24	PRIOR APPLICATION NUMBER: 60/047,612	24	PRIOR FILING DATE: 1997-05-23	24	PRIOR FILING DATE: 1997-08-22	24	PRIOR APPLICATION NUMBER: 60/056,874
25	PRIOR FILING DATE: 1997-05-23	25	PRIOR APPLICATION NUMBER: 60/047,632	25	PRIOR FILING DATE: 1997-08-22	25	PRIOR APPLICATION NUMBER: 60/056,874
26	PRIOR FILING DATE: 1997-05-23	26	PRIOR APPLICATION NUMBER: 60/047,601	26	PRIOR FILING DATE: 1997-08-22	26	PRIOR APPLICATION NUMBER: 60/056,874
27	PRIOR FILING DATE: 1997-05-23	27	PRIOR APPLICATION NUMBER: 60/043,580	27	PRIOR FILING DATE: 1997-08-22	27	PRIOR APPLICATION NUMBER: 60/056,874
28	PRIOR APPLICATION NUMBER: 60/043,580	28	PRIOR FILING DATE: 1997-04-11	28	PRIOR FILING DATE: 1997-08-22	28	PRIOR APPLICATION NUMBER: 60/056,874
29	PRIOR FILING DATE: 1997-04-11	29	PRIOR APPLICATION NUMBER: 60/043,568	29	PRIOR FILING DATE: 1997-08-22	29	PRIOR APPLICATION NUMBER: 60/056,874
30	PRIOR FILING DATE: 1997-04-11	30	PRIOR APPLICATION NUMBER: 60/043,314	30	PRIOR FILING DATE: 1997-08-22	30	PRIOR APPLICATION NUMBER: 60/056,874
31	PRIOR FILING DATE: 1997-04-11	31	PRIOR APPLICATION NUMBER: 60/043,569	31	PRIOR FILING DATE: 1997-08-22	31	PRIOR APPLICATION NUMBER: 60/056,874
32	PRIOR FILING DATE: 1997-04-11	32	PRIOR APPLICATION NUMBER: 60/043,311	32	PRIOR FILING DATE: 1997-08-22	32	PRIOR APPLICATION NUMBER: 60/056,874
33	PRIOR APPLICATION NUMBER: 60/043,311	33	PRIOR FILING DATE: 1997-04-11	33	PRIOR FILING DATE: 1997-08-22	33	PRIOR APPLICATION NUMBER: 60/056,874
34	PRIOR FILING DATE: 1997-04-11	34	PRIOR APPLICATION NUMBER: 60/043,671	34	PRIOR FILING DATE: 1997-08-22	34	PRIOR APPLICATION NUMBER: 60/056,874
35	PRIOR FILING DATE: 1997-04-11	35	PRIOR APPLICATION NUMBER: 60/043,674	35	PRIOR FILING DATE: 1997-08-22	35	PRIOR APPLICATION NUMBER: 60/056,874

```

; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,876
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,881
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,909
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,875
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,862
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,887
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,908
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/048,964
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/057,650
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/056,884
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/057,669
; PRIOR FILING DATE: 1997-09-05

```

Query Match 6.9%; Score 220; DB 10; Length 339;

Best Local Similarity 22.0%; Pred. No. 2.2e-13;
Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

```

OY 38 GEGRNPLEVQMSQVLVPCHTNNGELSVGQLKMDITACISARHAG-----CPCVT 91
DB 3 GPDVETPSAIGICRIMRDPDANVAGNVHGTLKMEBAGAIISTRHNSONGEVCVAL 62
OY 92 ASMDIYEHTSVOGVNIRAKVRAFNSMEVGIQVASEDLCEKQNNVCALATFYA 151
DB 63 ARVENTDPLSPMCIGEVAVHSAEITYTSKSHVEQVNVNMEIILGAKKLTNKATLWYVP 122
OY 152 RREITKVKLKQITP-----RTEEE-----KMEHSVAAERRRRRLVADTICKDLANCAIQD 203
DB 123 LSLKNVDKVLVEVPVAVSRXEQEEGRKRYEAKLERMETKRNQDIDVQPVLN----- 175
OY 204 LESRDCSRNVPAREKT-----RVESVELVLPFHANHOQNTFGQIMAMNEVATITASRLCR 259
DB 176 -----PEPNTVSYSSQSLIHLVGPSCDTLGFVHGVTMKLMDVAGIVARHCK 225
OY 260 AHPTLKAIEMFHRGPGSVQGRDLVLKAIYNNAFKSHMEVGVCEAVRQEAETHRRH-INS 318
DB 226 TMTVTASVDAINFHDKIRKGCVITISGRMTFTSNKSMELVLDADPVVDSSQKRYAAS 285
OY 319 AFMTFVLDDADDQPLLPMIRPOPGDGERRYRPARSAR 355
DB 286 AFTTYVLSIQEGRSLPVPQVLVPEDEKRRFEKGKR 322

```

```

RESULT 13
US-10-164-861-488
; Sequence 488, Application US/10164861
; Publication No. US20030225248A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/10/164,861
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US/09/149,476
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/04493
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 757
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 488
; LENGTH: 339

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (142)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (330)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (335)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (336)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (339)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals stop translation
US-10-164-861-488

```

Query Match 6.9%; Score 220; DB 12; Length 339;

Best Local Similarity 22.0%; Pred. No. 2.2e-13;
Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

```

OY 38 GEGRNPLEVQMSQVLVPCHTNNGELSVGQLKMDITACISARHAG-----CPCVT 91
DB 3 GPDVETPSAIGICRIMRDPDANVAGNVHGTLKMEBAGAIISTRHNSONGEVCVAL 62
OY 92 ASMDIYEHTSVOGVNIRAKVRAFNSMEVGIQVASEDLCEKQNNVCALATFYA 151
DB 63 ARVENTDPLSPMCIGEVAVHSAEITYTSKSHVEQVNVNMEIILGAKKLTNKATLWYVP 122
OY 152 RREITKVKLKQITP-----RTEEE-----KMEHSVAAERRRRRLVADTICKDLANCAIQD 203
DB 123 LSLKNVDKVLVEVPVAVSRXEQEEGRKRYEAKLERMETKRNQDIDVQPVLN----- 175
OY 204 LESRDCSRNVPAREKT-----RVESVELVLPFHANHOQNTFGQIMAMNEVATITASRLCR 259
DB 176 -----PEPNTVSYSSQSLIHLVGPSCDTLGFVHGVTMKLMDVAGIVARHCK 225
OY 260 AHPTLKAIEMFHRGPGSVQGRDLVLKAIYNNAFKSHMEVGVCEAVRQEAETHRRH-INS 318
DB 226 TMTVTASVDAINFHDKIRKGCVITISGRMTFTSNKSMELVLDADPVVDSSQKRYAAS 285
OY 319 AFMTFVLDDADDQPLLPMIRPOPGDGERRYRPARSAR 355
DB 286 AFTTYVLSIQEGRSLPVPQVLVPEDEKRRFEKGKR 322

```

```

RESULT 14
US-10-094-749-2967
; Sequence 2967, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU

```

```

; APPLICANT: OTSUKA, MOTOUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2967
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2967

```

Query Match 6.9%; Score 219; DB 15; Length 370;

Best Local Similarity 22.0%; Pred. No. 3.2e-13;

Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

```

QY 38 GEGYRNPTVEVQMSQVLVPCHTNORGEISVQGLTKWIDTTCLSAERHAG-----CPQVT 91
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 35 GPDVETPSAIOICRIMRDPDANVAGNVHGGTILKMIIGAGAITSTRHONSONGERCVAAL 94
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 ASADDIYEHTISVGVNITKAKVNRAFNSSMEVGIQVASEDLCEKQMNVCALATFVA 151
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 95 ARVERDTEFLPMCGEVAHVASETTYSKHSVEYQVVMSENILTGAKKLTNKTATLWYVP 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 RRRTTKYKLOIRP-----RTEERKMEHSVAERRMRLVY--ADITKOLLANCAIQGD 203
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 155 LSLKNVDKVLVEIPVNVYSROEGBEGRKRYEAQKLEMETKMRNGDVIQPVLN----- 207
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 204 LESRDCSRMVAEKT-----RVESVELVLPFHANQGTFGSQIMAMENVAITIASRLCR 259
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 208 -----PEPTVYSQSSLIHLVGPDSCTIHLGVHGGVTKMLDEVAQIVAAHCK 257
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 AHPTLKAIEHFHFGPSQVDRILVKAIVNAPFQSMVEGVCEVAEQEAETRRH-INS 318
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 258 TNITVTSVDAINFHDKIRKGCVITISGRMTFTSNKSKSEIVLVADADPVVDSQKRYAAS 317
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 319 AEMTFVLVDADDQPLLPIRPGDGERRYREASAR 335
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 318 APTTVSLSGRSLVPQVLPETDEKXRFEBGKR 354
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

```

US-10-282-122A-55313
; Sequence 55313, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

```

```

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55313
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-10-282-122A-55313

```

Query Match 6.5%; Score 207.5; DB 12; Length 159;

Best Local Similarity 33.3%; Pred. No. 1.2e-12;

Matches 46; Conservative 30; Mismatches 55; Indels 7; Gaps 2;

```

QY 226 VLPFHANQGTGGQIMAMENVAITIASRLCRPHFTLKAIEHFHFGPSQVDRILVK 285
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 24 IFPDINTNNTTIGGLIMSLDLRLALVAERHCESICVTALVDARFYAPAYMGENTLIC 83
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 AIYVNAFKHSMVEGVCEA---YRGEAETRRHINSFMTFFVLVDADDQQLPWIROP 342
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 84 ASVNRSMRISLSEGVAKMAENITKOE---ORHITSAYFTFVAVDENNSPVEIPELVET 139
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 GDGERRYREASARKIRL 360
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 QEIRRFHEADQRRASRL 157
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: August 12, 2004, 15:51:19

Job time : 50 secs

Blank Sheet

OM protein - protein search, using sw model

4559.989 Million cell updates/sec

Sequence: 1 MIQNVGNHLRLRGSLASVFSNR.....TLKAIPVAKINSRFGYLQDT 607

Gapor 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1:  sp_archea:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mbc:*
8:  sp_organelle:*
9:  sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	2458	77.3	579	11	Q8CA16	Q8CA16	O6CA16 mus musculi
2	2421	76.1	561	11	Q8OT19	Q8OT19	mus musculi
3	255	8.0	339	16	Q8NM17	Q8NM17	colynebactere
4	233	7.3	335	16	Q8FM15	Q8FM15	colynebactere
5	228	7.2	165	17	Q97B13	Q97B13	thermoplasma
6	224.5	7.1	276	16	Q8U912	Q8U912	agrobacteri
7	217.5	6.8	165	17	Q9HK19	Q9HK19	thermoplasma
8	216.5	6.8	157	16	Q9KE11	Q9KE11	bacillus ha
9	216	6.8	162	16	Q9KA15	Q9KA15	bacillus ha
10	211.5	6.6	156	16	Q82416	Q82416	chlamydophi
11	207.5	6.5	161	16	Q87127	Q87127	pseudomonas
12	204	6.4	167	16	Q92TC6	Q92TC6	rhizobium m
13	204	6.4	318	17	Q97WD9	Q97WD9	sulfolobus
14	203	6.4	168	16	Q81R11	Q81R11	bacillus an
15	201	6.3	188	16	Q9HT14	Q9HT14	pseudomonas
16	201	6.3	207	16	Q82DP3	Q82DP3	streptomyce

17	199	6.3	313	17	Q26Y86
18	198.5	6.2	153	16	Q7VFM3
19	198.5	6.2	354	2	Q9RHY3
20	197.5	6.2	163	16	Q83C60
21	197.5	6.2	197	16	Q93353
22	197	6.2	171	16	Q81EB4
23	191.5	6.0	161	16	Q88C27
24	189	5.9	315	2	Q33537
25	188	5.9	171	16	Q8EM17
26	185	5.8	176	16	Q8CR75
27	184	5.8	438	10	Q49337
28	180.5	5.7	166	16	Q91EB9
29	179.5	5.6	163	16	Q7MB80
30	179.5	5.6	164	16	Q8XB0
31	179.5	5.6	171	16	Q7WCB9
32	179	5.6	160	16	Q9JTP2
33	179	5.6	382	5	Q9JTP2
34	179	5.6	160	16	Q9JTP2
35	174.5	5.5	365	17	Q9Y939
36	174	5.5	172	16	Q92A62
37	173	5.4	172	16	Q8YV99
38	172	5.4	133	16	Q8EP27
39	171.5	5.4	130	16	Q8LTM8
40	171	5.4	133	16	Q8EPB3
41	170.5	5.4	176	16	Q931N3
42	170.5	5.4	176	16	Q931N3
43	170.5	5.4	307	17	Q975X5
44	170.5	5.4	472	5	Q9Y9M4
45	166	5.2	178	16	Q831O6

ALIGNMENTS

Q9676b sulfolobus
Q97vm3 helicobact
Q9rh3c corynebacte
Q9h360 coriella b
Q93j53 streptomyces
Q98ie4 bacillus c
Q98e27 pseudomonas
Q35517 rhodococcus
Q9em17 oceanobac
Q9acv5 staphylococ
Q94937 arabadipos
Q94ie9 pseudomonas
Q97x80 bordetella
Q98xb0 ralistonia s
Q97we3 bordetella
Q94j0o neisseria m
Q94tp2 neisseria n
Q9ywf5 aerophylla
Q9yfl9 aerophylla
Q92ae2 listeria m
Q985v9 listeria m
Q983t7 xanthomonas
Q98fj8 bacillus an
Q98fb3 xanthomonas
Q93ic3 strephylOCO
Q93ic0 strephylOCO
Q975x5 sulfolobus
Q9ywf4 drosophila
Q93iq6 enterococcus

RESULT 1

ID	Q8CAL6	PRELIMINARY;	PRT;	579 AA
AC	Q8CAI6			

AC Q8CAL6;

DT	01-MAR-2003	(TREMBlrel. 23, Created
DT	01-MAR-2003	(TREMBlrel. 23, last 20

DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)
----	---

DE BROWN Inducible thioesterase 2.
GN THEA.

Mus musculus (Mouse).

Mammalia; Eutheria

```

OX      NCBI_TaxID=10090;
RN      [1]

```

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Hypothalamus
RX MEDLINE=22354683; Pubmed-12466851.

RA The FANTOM Consortium,

RT Analysis of the mouse transcriptome based on functional annotation of the mouse genome Exploration Research Group Phase I & II Team;

RT	60, 770 full-length cDNAs.";
RT	Nature 400:553-558 (1998).

EMBL; AK038546; BAC30036.1; -

DR MGD; MGI:1913736; Thea.
DR GO: GO:0002824; F. scholasticus

DR InterPro; IPR002913; START.

```
DR      InterPro; IPR0066683; Thioestr_supf
DR      Pfam: PF03061; 4HBT: 3
```

DR PFam; PF01852; START; 1.

```
DR      PROSITE: PS50848: START: 1.
DR      SPEAK1; SM00234; START; 1.
DR
```

SEQUENCE 579 AA; 65672 MW; C832529379C11F5A CRC64;

Query Match 77.3%: Score 2458: DB 11: Length 579

Best Local Similarity 81.6%; Pred. No. 8.4e-205;
Matches 479, Consensus 35

..... 33; Indels 20; Gaps 6

16 VFSNRTSRKSAI--RAGNDSAMADGEGYRNPTVEQMSQLVLPCHTNQRGELSVGQLLKI 73

Db 1 MFSNRTSRKSI SHPESGDDPTMAEGEGYRNPTVEQMSQVLVPCHTNHRGELSIGQLLKW1 60
 QY 74 DTTAGLSAERHACPCVTASMDIYFPHITISVGVVNIKAKVNAFNSMEVGIQVASED 133
 Db 61 DTTAGLSAERHACPCVTASMDIYFPHITISVGVVNIKAKVNAFNSMEVGIQVASED 120
 QY 134 LGSCKQNVCKALATFYAREITVVKLKOITPRTEEEKMEHVAERRRRLVADTIKD 193
 Db 121 LGSCKQNVCKALATFYAREITVVKLKOITPRTEEEKMEHVAERRRRLVADTIKD 180
 QY 194 LLSNCAIQGLSRRDCSRMTPAEKTRVESVELVLPphanhQNTFGQIMAMENVAITA 253
 Db 181 LLSNCAIQGLSRRDCSRMTPAEKTRVESVELVLPphanhQNTFGQIMAMENVAITA 239
 QY 254 ASRLGRAHPTLKAIEHFHFRGSGVGRDLVKAIVNNAFKHSMGVGVCEAYROBAETHR 313
 Db 240 ASRLGRAHPTLKAIEHFHFRGSGVGRDLVKAIVNNAFKHSMGVGVCEAYROBAETHR 299
 QY 314 RHINSAMTFVVLDAADQPOLFWIRPQPGDGERRYEASARKKIRLDRKXIVSCQTEV 373
 Db 300 RHINSAMTFVVLDAADQPOLFWIRPQPGDGERRYEASARKKIRLDRKXIVSCQTEV 359
 QY 374 PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLBEDKFLSFHMEVVA 433
 Db 360 ALSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLBEDKFLSFHMEVVA 418
 QY 434 VDAQAQFLSLDLRORPEMDKHYSVELVQVDEDDAIYHTSPALGHTKPODFVILLAS 493
 Db 419 VDAQAQFLSLDLRORPEMDKHYSVELVQVDEDDAIYHTSPALGHTKPODFVILLAS 478
 QY 494 RRKPCDNDPPYIALRSTVLTPTHEPTRYRGELTCSGCLMREBDQTLTKCWNVSLTE 553
 Db 479 RRKPCDNDPPYIALRSTVLTPTHEPTRYRGELTCSGCLMREBDQTLTKCWNVSLTE 533
 QY 554 LVSAGFYWGLESRSKGRSPDNGKLAGHSLTLKAIIPAKINSR 600
 Db 534 -QATPGFLNY-VTTNVSGLSSEFYN-----TFKACESFLDNR 569

RESULT 2
 ID 080TT9 PRELIMINARY; PRT: 561 AA.
 AC 080TT9;
 DT 01-JUN-2003 (TReMBLrel. 24, Created)
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE MKIAA0707 protein (Fragment).
 GN MKIAA0707.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Ynasa S.,
 Nakajima D., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT II. The complete nucleotide sequences of 400 mouse KIAA-homologues
 RT randomly sampled from size-fractionated libraries.";
 RL DNA Res. 10:35-48(2003).
 RL EMBL; AK122350; BAC65632.1; -;
 DR GO; GO:0003824; F: catalytic activity; IEA.
 DR InterPro; IPR002913; START.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 2.
 DR Pfam; PF01852; START; 1.
 DR SMART; SM00234; START; 1.
 DR PROSITE; PSS0948; START; 1.
 FT NON_TER 1 1
 SQ SEQUENCE 561 AA; 63654 MW; 6D7183C48F103EAA CRC64;

Query Match 76.1%; Score 2421; DB 11; Length 561;
 Best Local Similarity 82.9%; Pred. No. 1.3e-201;
 Matches 469; Conservative 33; Mismatches 47; Indels 18; Gaps 5;

QY 35 MADGEGYRNPTVEQMSQVLVPCHTNHRGELSIGQLLKW1 94
 Db 4 MADGEGYRNPTVEQMSQVLVPCHTNHRGELSIGQLLKW1 63
 QY 95 DDYFEHTISVGVVNIKAKVNAFNSMEVGIQVASEDLSCKQNVCKALATFYARE 154
 Db 64 DDYFEHTISVGVVNIKAKVNAFNSMEVGIQVASEDLSCKQNVCKALATFYARE 123
 QY 155 ITKVKLKOITPRTEEEKMEHVAERRRRLVADTIKDILANCAIQGLSRRDCSRMTP 214
 Db 124 LSKYKLMQVLPLEEKTEHGAERRRRLVADTIKDILTHCVIIDDLD-KDCSNMVP 182
 QY 215 AEKTRVESVELVLPphanhQNTFGQIMAMENVAITAASRLGRAHPTLKAIEHFHFRG 274
 Db 183 AEKTRVESVELVLPphanhQNTFGQIMAMENVAITAASRLGRAHPTLKAIEHFHFRG 242
 QY 275 PSQVGRDLVKAIVNNAFKHSMGVGVCEAYROBAETHRRLHNSAFMTFVLDAADQPOL 334
 Db 243 PSQVGRDLVKAIVNNAFKHSMGVGVCEAYROBAETHRRLHNSAFMTFVLDAADQPOL 302
 QY 335 LPWIRPQPGDGERRYEASARKKIRLDRKXIVSCQTEVPLSPWMDPSNOVYLSYNNVSS 394
 Db 303 LPWIRPQPGDGERRYEASARKKIRLDRKXIVSCQTEVPLSPWMDPSNOVYLSYNNVSS 362
 QY 395 LKMTLAKDNWVLSSEISQVRLYTLBEDKFLSFHMEVVAHVAQAQFLSLDLRORPEMDK 454
 Db 363 LKMTLAKDNWVLSSEISQVRLYTLBEDKFLSFHMEVVAHVAQAQFLSLDLRORPEMDK 421
 QY 455 HYSVELVQVDEDDAIYHTSPALGHTKPODFVILLASRRKPCDNDPPYIALRSTVLT 514
 Db 422 HYSVELVQVDEDDAIYHTSPALGHTKPODFVILLASRRKPCDNDPPYIALRSTVLT 481
 QY 515 THRETPRYRGELTCSGCLMREBDQTLTKCWNVSLTELVASAGFYWGLESRSKGRS 574
 Db 482 THRETPRYRGELTCSGCLMREBDQTLTKCWNVSLTELVASAGFYWGLESRSKGRS 534
 QY 575 DGMNGKLAGHSLTLKAIIPAKINSR 600
 Db 535 EFYN-----TEKACESFLDNR 551

RESULT 3
 ID 08NM17 PRELIMINARY; PRT: 339 AA.
 AC 08NM17;
 DT 01-OCT-2002 (TReMBLrel. 22, Created)
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE Acyl-CoA hydrolase.
 GN CGL2586.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 OC NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF005282; BAB99979.1; -;
 DR GO; GO:0016787; F: hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 2.
 DR Hydrolase; Complete proteome.
 KW SEQUENCE 339 AA; 37521 MW; A2305D953D604AFD CRC64;

Query Match 8.0%; Score 255; DB 16; Length 339;

RESULT 4	ID	Q8FMMS	PRELIMINARY;	PRT;	335 AA.
Q8FMMS	AC	Q8FMMS;			
DT	01-MAR-2003	(T-EMBLrel. 23, Created)			
DT	01-MAR-2003	(T-EMBLrel. 23, Last sequence update)			
DT	01-OCT-2003	(T-EMBLrel. 25, Last annotation update)			
DE	Conserved hypothetical protein.				
GN	CE2478.				
OS	Corynebacterium efficiens.				
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OC	Corynebacterineae; Corynebacteriaceae; Corynebacterium.				
OX	NCBI_TaxID=152794;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;				
RA	Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,				
RA	Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,				
RT	Usuda Y., Sugimoto S.;				
RL	"The entire genomic sequence of Corynebacterium efficiens YS-314."				
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; A0005222; BAC19288.1; -				
DR	GO; GO:0003824; F:catalytic activity; IEA.				
DR	InterPro; IPR006683; Thioestr_supt.				
DR	Pfam; PF03061; 4HPT; 2.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 335 AA; 36744 MW; 6985934FF1F837D76 CRC64;				

		7.3%	Score 233;	DB 16;	Length 335;	
		Best Local Similarity	23.7%;	Pred. No. 1,2e-11;		
		Matches	79;	Conservative	55;	Mismatches 178; Indels 22; Gaps 6
QY	35	MADGGRNP--TEVMSQLVPCHTNRGELSV--GOLLKVIDPTACTASRHAGPCPV	90			
Db	1	MSDQQTASPSPDVTIRFLAFTVDVLMAAGSHGVGGRLBMDIKAAVACATQWSGTVCY	60			
QY	91	TASMDIYEFTISVGCVVINAKAKNRAFNSMMEGIQVASDDLCEKQNVCKALATFV	150			
Db	61	TAYGHHTFPITPSGHAVVEYSRKIAMTRGSMMHVNENVLADPPDGFRTRACDLIVTFV	120			
QY	151	ARREIT--KYVKLOITPRTEEKEKHSVAERRRMLTYADTIKILANCAIOGLESRD	208			
Db	121	AKDATATGTOEHPFVFPEKTAEERVLQSALSRIIDR-----ALEAMEQOT	167			
QY	209	CSRVRPAEKRRVESVELVLPFHANHGQTFEGQIMAMKENNAITIASLCRAHPTLKAI E	268			

ID	Q97B23	PRELIMINARY;	PRT;	165 AA.
AC	Q97B23;			
DT	01-OCT-2001 (TREMBlrel. 18, Created)			
DT	01-OCT-2001 (TREMBlrel. 18, Last sequence update)			
DT	01-JUN-2003 (TREMBlrel. 24, Last annotation update)			
DE	Acyl coenzyme A thioester hydrolase.			
GN	TV0636 OR TVG0629193.			
OS	Thermoplasma volcanicum.			
OC	Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;			
OC	Thermoplasmataceae; Thermoplasma.			
OK	NCBI_TaxID=50339;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RX	STRAIN=GSST / DSM 4299 / UCM 9571;			
RX	MEDLINE=20570466; PubMed=11121031;			
RA	Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,			
RA	Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,			
RA	Nunoshiba T., Yamamoto Y., Aramaki H., Nakino K., Suzuki M.;			
RT	"Archaeal adaptation to higher temperatures revealed by genomic			
RT	sequence of Thermoplasma volcanicum."			
RL	Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262 (2000).			
RL	EMBL: AP000993; BAB59778.1;			
DR	GO:GO:0016787; F:hydrolase activity; IEA.			
DR	InterPro: IPR006683; Thioestr_supf.			
DR	Pfam: PF03061; 4HBT, 1.			
KW	Hydrolase; Complete proteome.			
SO	SEQUENCE 165 AA; 19195 MW; 5392888A47B2769D CRC64;			

Query Match	7.2%	Score 228	DB 17	Length 165
Best Local Similarity	34.6%	Pred No. 1.2e-11		
Matches	55	Conservative 35	Mismatches 61	Indels 8
				Gaps 5
QY	214	PAE-KTRVESVELVLP.PPHANQGNFTFGQIAMMENVNTVITIASRLCPRAHPTLKAIEMHFH	272	
DB	6	PSESETFVE--RMVLPEMDMIEYLYLGGRLMEMIDNCASIVATKRCRRRTVIGSIDSLFF	63	
QY	273	RGPSQVGRVLVKALVNNAPFHSNEVGCVAAYROEATH--RRHINSAPMTFVVLADADQ	331	
DB	64	LPLPHLGSMVLHGVINNTTSTWIE--DIVIKQEGLTGIRRYATKAYLVVAIDADGR	121	
QY	332	FQLPWIRPQPGDGRRRRYEASAKKILDRKCVYSCQ	370	
DB	122	PTEIRPOLIEPTDDKKRYAEAKAEER--RKRLETTIQ	158	

RESULT 6

ID	Q8U9A2	PRELIMINARY	FRT	276 AA.
AC	Q8U9A2			
DT	01-JUN-2002	(TremblereL. 21, Created)		
DT	01-JUN-2002	(TremblereL. 21, Last sequence update)		
DT	01-JUN-2003	(TremblereL. 24, Last annotation update)		
DE	ACY1-COA	hydrolase.		
GN	ATU3826	OR AGR L 2016.		
OS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.			
NCBI	TaxID=176299;			
XX	[1]			
NN				

RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Boye D. Sr.,
 RA Chapman P., Clendinning U., Deatherage G., Gillet W., Grant C.,
 RA Kutyavin T., Levy R., Li M.-U., McClelland E., Palmieri A., Gordon D.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero F., Perry M.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley P., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Neeter E.W.,
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 C58.";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Quirillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmlet K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
 RA Mollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Garsen J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Sister S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328(2001).
 DR EMBL; AF009314; ALA4636.1; ALT_INIT.
 DR EMBL; AF008300; AAK89583.1; -.
 DR PIR; AF3027; AF3027.
 DR PIR; E98257; E98257.
 DR GO; GO:0016787; F:Hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 2.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 276 AA; 29993 MW; 42031EB95951E76E CRC64;

Query Match 7.1%; Score 224.5; DB 16; Length 276;
 Best Local Similarity 25.2%; Pred. No. 5.1e-11;
 Matches 72; Conservative 42; Mismatches 127; Indels 45; Gaps 5;

QY 48 QMSQVLRCHNORELSVGLKWIPTTACLSARHNGCCPVTSASMDIYEHTISVQ 107
 DB 18 RLIDIVFEDTNGHGTGEGTGLAMDRVAFIAIRFGPTPTVTSACRIDRQARIGH 77
 QY 108 VVNIKAVRAAFNSMEVGIQVASEDLCEKQWVCKALATFVARREITKVKLKITPRT 167
 DB 78 IVEFTAPRPVKGRRLTYVEVWVAFTIIGRQHTCRGIFHVA-----IPRG 125
 QY 168 EEEKMEHSVAAERRMRLVYADTIDQLANCAIQDLESRDCSRWPAKTEVESVELYL 227
 DB 126 ED-----ASVYLPBELTET-----PDSDAVTWELVF 155
 QY 228 PEHANHGQTFGGQIMAMMENVAATIASRLCRAPHTLTKAIEMFHFGPQVGDRLTKAI 287
 DB 156 PQOASAGMFGGEALAYTKAFVAASVYCGKLVILASERKIDPRARALEIGIYEAQKH 215
 QY 288 VVNAFEGHSMGVGVCEAYROEAEYTHRRHIN-SAFMTFVVLDDDDP 332
 DB 216 VERVGRSSMSIG-TKLWSENILLTGRHITATGFTMVAVDKDRP 259

RESULT 7
 Q9HKJ9 PRELIMINARY; PRT; 165 AA.
 AC Q9HKJ9;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Acyl-CoA hydrolase related protein.
 GN TA0599.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;

OC Thermoplasmataceae; Thermoplasma.
 CX NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Rupp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 acidophilum.";
 RL Nature 407:508-513(2000).
 DR EMBL; AL45064; CAC11738.1; -.
 DR GO; GO:0016787; F:Hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 1.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 165 AA; 19233 MW; 28F1614521FPA4DA CRC64;

Query Match 6.8%; Score 217.5; DB 17; Length 165;
 Best Local Similarity 32.3%; Pred. No. 9.6e-11;
 Matches 50; Conservative 38; Mismatches 60; Indels 7; Gaps 4;

QY 217 KTRVESVELVPPYAHNGQTFGGQIMAMMENVAATIASRLCRAPHTLTKAIEMFHFGPS 276
 DB 10 ETVYE--RMVLPEDMDIYELYGRLEMWIDNCASIVATKCRKRTGTSIDSLFFLLPI 67
 QY 277 QVGRIVLTKAVNNAFKSMGVCEAYROEAEYTHRRHINSAPMTFVVLDDDDPQL 335
 DB 68 HLGDMVILHGYINTYTSMTWEI-IDVYKEGLTGIRRYATKALTYVAIDSDGRPREI 125
 QY 336 FWIRPQDGERRYEASARKKIRIDRYIVSCQ 370
 DB 126 PQIVETDEKRRYQDAEKRAER--RKRLSINQ 158

RESULT 8
 Q9KEQ1 PRELIMINARY; PRT; 157 AA.
 AC Q9KEQ1;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Acyl-CoA hydrolase.
 GN BH0798.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AF001509; BAB04517.1; -.
 DR PIR; F83749; F83749.
 DR GO; GO:0016787; F:Hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 1.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 157 AA; 17328 MW; 8639B00827EEBA3A CRC64;

Query Match 6.8%; Score 216.5; DB 16; Length 157;
 Best Local Similarity 32.0%; Pred. No. 1.1e-10;
 Matches 48; Conservative 32; Mismatches 67; Indels 3; Gaps 2;

QY 214 PAEKTVESVELVPPYAHNGQTFGGQIMAMMENVAATIASRLCRAPHTLTKAIEMFHFR 273
 DB 6 PVERSRITQRLVLPDPVNLGTFGSKVLAIDEIALTKMHANSAVVTASIDSDVDFK 65

QY	274	GPQGVGRILVKAIVNNAFHGSHVEGVCVAHYQAEFTRRHINS-APMTFVLADADDP	332
Db	66	SSATVGDALHEGFTVHTGTSMH--VYRVHSHNNLLTGRITLTSTESFLTMVAVDSEGKP	123
QY	333	QLPWIRPQDGERRYREASARKKIRLDR	362
Db	124	KVPVQVEPQTEERKRLYETRPARKENRKR	153
RESULT 9			
ID	Q9KAI5	PRELIMINARY;	PRT; 162 AA.
AC	Q9KAI5;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, last sequence update)		
DE	01-JUN-2003 (TrEMBLrel. 24, last annotation update)		
DE	Acyl-CoA thioester hydrolase.		
GN	BH2302.		
OS	Bacillus halodurans.		
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
OX	NCBI_TaxID=86665;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C-125 / JCM 9153;		
RX	MEDLINE=20512582; PubMed=11058132;		
RA	Takemi H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,		
RA	Fuji F., Hirata C., Nakamura Y., Ogasawara N., Kuhara S.,		
RA	Horikoshi K.;		
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus		
RT	halodurans and genomic sequence comparison with Bacillus subtilis."		
RL	Nucleic Acids Res. 28:4317-4331(2000).		
DR	EMBL, AP001515; BAB06021.1; --		
DR	PIR, F83937; F83937.		
DR	GO; GO:0016787; F:hydrolase activity; IEA.		
DR	InterPro; IPR006683; Thioestr_supf.		
DR	Pfam; PF03061; 4HBT. 1.		
KV	Hydrolase; Complete proteome.		
SEQ	SEQUENCE 162 AA; 18443 MW; 5FB37A3B1B61CC CRC64;		
Query Match 6.8%; Score 216; DB 16; Length 162;			
Best Local Similarity 35.1%; Pred. No. 1,3e-10;			
Matches 53; Conservative 31; Mismatches 63; Indels 4; Gaps 3;			
QY	214	PAEKTRVSVSLVLPHPANHOQNTFGQIMAMENVATIASRLCPRAPTLKAIEMFHR	273
Db	7	PSESLITWT-DIVLPEDTNVHGTLFGGVMAVYVKVASIAMRCKRKQVYVASSDSLDFI	65
QY	274	GPSQGVGRILVKAIVNNAFHGSHVEGVCVAHYQAEFTRRHIN-SAFMTFVLADADDP	332
Db	66	SPITGEALICEGVLYTRKTSMEVFAVEA--EDLTGERRLTATSYLTFVALINPDGKP	123
QY	333	QLPWIRPQDGERRYREASARKKIRLDR	363
Db	124	AEIPPVVPTTEERKRLYETRPARKENRKR	154
RESULT 10			
ID	Q824Q6	PRELIMINARY;	PRT; 156 AA.
AC	Q824Q6;		
DT	01-JUN-2003 (TrEMBLrel. 24, Created)		
DT	01-JUN-2003 (TrEMBLrel. 24, last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, last annotation update)		
DE	Cytosolic acyl-CoA thioester hydrolase family protein.		
GN	CAA00086.		
OS	Chlamydomonadales.		
OC	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomonadales.		
OX	NCBI_TaxID=83557;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=GPIC;		
RX	MEDLINE=22569155; PubMed=12682364;		
RA	Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,		

	R	RA	Haidelberg J., Holzapfel E., Khouri H., Fedorova N.B., Carty H.A., Mayan L.A., Haft D.H., Peterson J., Beaman M.J., White O., Salzberg S.L., Hsia R.-C., McClary G., Rank R.G., Bavoil P.M., Fraser C.M.; "Genome sequence of Chlamydomonas reinhardtii examining the role of niche-specific genes in the evolution of the Chlorophyta"; Nucleic Acids Res. 31:213-214(2003).
	RL	EMLL_AE016894; AAP04038.1; --.	
	DR	TIGR; CCAG00086; --.	
	DR	GO; GO:0016787; F:hydrolyase activity; IEA.	
	DR	InterPro; IPR006683; Thioestr_supf.	
	DR	Pfam; PF03061; 4HBT; 1.	
	KM	Hydrolase; Complete proteome.	
SQ	SEQUENCE	156 AA; 17908 MW; CFA49E2FB6EFDD94 CRC64;	
	Query Match	6.6%; Score 211.5; DB 16; Length 156;	
	Best Local Similarity	34.5%; Pred. No. 2_9e-10;	
Matches	48; Conservative	28; Mismatches	56; Indels 7; Gaps 2
QY	226	VLPFHANHOAGNTFGGCIAMMENAVATIAASRLCAHPHTKALIMFHRGPSQVGDRLYLK	285
DB	19	IPEPDINTNNTPTFGGILMSLDRLALVALVERHTEGICVTAFDALRFYPAPWGENULICK	78
QY	286	AIVNNAFKHSMEVGVCVA---YRCGEFTRHRTINSAEPTFFVLVDADDOPQLIPWTIRPQP	342
DB	79	AAVNRTWKTSLSGVAKVMENIYIKOR---RRHITSAYTTFVSVDENNAPPVHQVVPER	134
QY	343	GDEERRRYEARSAKKXIRLD	361
DB	135	PEKKRRYEDADQRQSRLN	153
RESULT	11		
ID	Q87TZ7	PRELIMINARY;	PRT; 161 AA.
AC	Q87TZ7;		
DT	01-JUN-2003 (TREMBLRel. 24, Created)		
DI	01-JUN-2003 (TREMBLRel. 24, Last sequence update)		
DE	Cytosolic long-chain acyl-CoA thioester hydrolase family protein.		
GN	PSPO5520.		
OS	Pseudomonas syringae (pv. tomato).		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
CC	Pseudomonadaceae; Pseudomonas.		
OX	NCBI_taxid=323;		
NB	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=DC3000:		
RC	Buell R., Jeardar V., Khouri H., Fedorova N., Tran B., Russell D., Berry K., Uteerbeck T., Van Aken S., Feldblum T., Gwin M., Dodson R., DeBooy R., Durkin A., Kolonay R., Madupu R., Daugherty S., Brinkac L., Beaman M., Haft D., Selengut J., Nelson W., Davidson T., White O., Fraser C., Collier A.; "Complete sequence of Pseudomonas syringae "; Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.		
RL	EMBL; AE016876; AAC08939.1; --.		
DR	TIGR; PSPT05520; --.		
DR	GO; GO:0016787; F:hydrolyase activity; IEA.		
DR	InterPro; IPR006683; Thioestr_supf.		
DR	Pfam; PF03061; 4HBT; 1.		
KM	Hydrolase; Complete proteome.		
SQ	SEQUENCE	161 AA; 18167 MW; B0842BFSD4542626 CRC64;	
	Query Match	6.5%; Score 207.5; DB 16; Length 161;	
	Best Local Similarity	34.0%; Pred. No. 6.8e-10;	
Matches	49; Conservative	30; Mismatches	58; Indels 7; Gaps 3;
QY	225	LVPFHANHOAGNTFGGCIAMMENAVATIAASRLCAHPHTKALIMFHRGPSQVGDRLYL	284
DB	13	IMPEDMANFSGNHGGLTKLYLDEVAVACAASRVAGRYYTLISDVQVFIREPVHGELVTF	72
XY	285	KAIYNNAFKHSMEVGVCV--EAVRGEFTRHRTINSAEPTFFVLVDADDOPQLIPWTIRPQP	342

RA Fraser C.M.;
 RT "The genome sequence of *Bacillus anthracis* Ames and comparison to
 RT closely related bacteria."
 RL Nature 423:81-86(2003).
 DR EMBL; AE017030; AAP2594.1; -.
 DR TIGR; BA2053; -.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 1.
 KM Hydrolase; Complete proteome.
 SQ SEQUENCE 168 AA; 18926 MW; 87478FDE8C4ACB34 CRC64;

Query Match 6.4%; Score 203; DB 16; Length 168;
 Best Local Similarity 35.0%; Pred. No. 1.8e-09;
 Matches 55; Conservative 27; Mismatches 71; Indels 4; Gaps 2;

QY 43 NPTVQMSQVLPCHTNGELSVGOLLKWIPTTACLSAERHAGCPCTASMDIYEHT 102
 DB 10 NESRVFMTSRVFPDNDHNTLPFGKTLAEMDMVATISATRHAKRECVTASMDWDFLHP 69
 QY 103 ISVQGVNIRAKVRAAFNSNEVGIOVASEDLCSKQNNVCALATFVA-FREITKYKLK 161
 DB 70 VRSDCVSYSEFVMTGRTSMEEVFKVAEDLISGEKRIATSPVTFVALSKENNPVPVP 129
 QY 162 QITPREEMEHSA--AERRMRLVADTIKDL 195
 DB 130 RVIPETBEKEHLRIAVLRABQRIHRAESKRVATLL 166

RESULT 15

Q9HTS4 PRELIMINARY; PRT; 188 AA.
 ID 09HTS4
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein PA5519.
 GN PA5519.
 OS *Pseudomonas aeruginosa*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino R., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Keizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 RT opportunistic pathogen."
 RT Nature 406:959-964 (2000).
 DR EMBL; AE004964; AAC08904.1; -.
 DR PIR; F82956; F82956.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 188 AA; 21215 MW; 8705C22F1C9F847E CRC64;

Query Match 6.3%; Score 201; DB 16; Length 188;
 Best Local Similarity 30.2%; Pred. No. 3.2e-09;
 Matches 54; Conservative 36; Mismatches 69; Indels 20; Gaps 3;

QY 13 IASVSNRKS-----RKSLRAANDSAMADGGRNPTVQMSQVLPCHTNGELSV 66
 DB 9 LPSSPASVSVSPARPERKSAMQPGNN-----QLSMTVLMTDPDMANFSGNVHG 55
 QY 67 GQLTKMIDTTACLSAERHAGCPCTASMDIYEHTISVQGVNIRAKVRAAFNSMEVG 126

DB 56 GTLLKYIDEVAVACASRYAGRVVTLASDVQVIFREPVAHVGEIVFLASVNYTGRTSMELG 115
 QY 127 IOVASEDLCSKQNNVCALATFVARREITK-VKLKQITPREEMEHSAERRRMR 184
 DB 116 VKVITENIREQSVHHTNSCFFTWALDDEKRPVAVRPLELETAERFAQAOQRQRQR 174

Search completed: August 12, 2004, 15:45:21
 Job time : 44 secs

Blank sheet


```
/ EARLIER APPLICATION NUMBER: 60/047,593
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,614
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/043,578
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,576
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/047,501
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/043,670
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/056,632
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,664
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,876
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,881
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,909
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,875
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,862
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,887
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,908
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/048,964
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/057,650
/ EARLIER FILING DATE: 1997-09-05
/ EARLIER APPLICATION NUMBER: 60/056,884
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/057,669
/ EARLIER FILING DATE: 1997-09-05
/ EARLIER APPLICATION NUMBER: 60/049,610
/ EARLIER FILING DATE: 1997-06-13
/ EARLIER APPLICATION NUMBER: 60/061,060
/ EARLIER FILING DATE: 1997-10-02
```

Query Match 6.9%; Score 220; DB 4; Length 339;

Best Local Similarity 22.0%; Pred. No. 6,4e-16; Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

```
QY 38 GEGYRNPTEYOMSLVLPCHTNGELSVGLKMTDTTCLSAERHAG-----GCVT 91
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3 GPDVEFSAIQIRKIRPDANVAGVHGTTIKMIEFAGAITSTRHCNSQNGERCVAAL 62
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 ASMDIYFEHTTSVGVGVNIKAVNRAFNSSMEVGIQVASEDLCSKQNNVCKALATFYA 151
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 63 ARVERTDIFSPMCI GEVAHVSABITYTSKSHVEVQVNVASENLTGAKKLTNATLWYV 122
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 RREITVYKUKQITP-----RTEER---KHEHSVAERRRRLVYATTKDLANCAIQSD 203
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 123 LSHKNDKLVLEVPVYVSRKEQEEGRKRYEACKLRMETKMNNGDIVPVLN----- 175
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 204 LBSRDCSRWVPAEKT-----RVESVELVLPFHANHCNTRGQIMAMMENAVATIASRLR 259
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 176 -----PENNTYSYQSLSIHLVGSDDCTLHGFVHGVTWKLMDVAGIVAAHCK 225
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 AHPTLKAIMFHFGRGSQVGDRLVLKAIYNNAFKHSMVEGVCEAVYROAEATERRH-INS 318
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 226 TIVIVTASVDALNFHDKIRKGCVTITISGRMTFTSNKSMETIEVLVDADPVVDSQKRYAAS 285
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 319 AEMTFVLADADQPOLLPWIRPQPGGGRYRARSAR 355
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 286 AFTTVISLSQEGSLFPQIVPETTEDEKRFEGKGR 322
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
US-09-252-991A-19094
/ Sequence 19094, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196,136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 31142
/ SEQ ID NO 19094
/ LENGTH: 202
/ TYPE: PRN
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19094
```

Query Match 6.3%; Score 201; DB 4; Length 202;

Best Local Similarity 30.2%; Pred. No. 3,7e-14; Matches 54; Conservative 36; Mismatches 69; Indels 20; Gaps 3;

```
QY 13 LASVSNRNTS-----RKSALRAGNDSAMADGEGYRNPTEYOMSLVLPCHTNGELSV 66
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 23 LPSEFASWBPARDDEKSMQNGNN-----QSMVLTMPDMANSGVNHG 69
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 GQLKMTDTTACLSAERHAGCPCTYASMDIYFEHTTSVGVGVNIKAVNRAFNSSMEVG 126
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 70 GTLKYIDEVAAYACASRYAGRYVYVTLSDQVTFREPHVGLVTFSLASVNTGRTSMEIG 129
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 127 IOVASEDLCSKQNNVCKALATFYAREITK-VLKQITRTTEKHEHSVAERRRMR 184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 130 VKVITENIRQSVRHTNSCFTFWALDEDERKPAVRLLETAEQKRFAQACQROQIR 188
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 4

US-09-328-352-5691

/ Sequence 5691, Application US/09328352

/ Patent No. 6562958

/ GENERAL INFORMATION:

/ APPLICANT: Gary L. Breton et al.

/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

/ FILE REFERENCE: GTC99-03PA

/ CURRENT APPLICATION NUMBER: US/09/328,352

/ CURRENT FILING DATE: 1999-06-04

/ NUMBER OF SEQ ID NOS: 8252

/ SEQ ID NO 5691

/ LENGTH: 172

/ TYPE: PRN

/ ORGANISM: Acinetobacter baumannii

US-09-328-352-5691

Query Match 6.0%; Score 191.5; DB 4; Length 172;

Best Local Similarity 32.6%; Pred. No. 3,3e-13; Matches 47; Conservative 26; Mismatches 70; Indels 1; Gaps 1;

```
QY 225 LVLPPHANHCNTRGQIMAMMENAVATIASRLCRAHPTLKALIMFHFGRSPQVGDRLVL 284
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 23 LMTPDMANFSQVHGTTIKLIDQVAYACASRYSGSYVTLSDKXNFKEPIYVGLVTF 82
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 KAIYNNAFKHSMVEGVCEAVYROAEATERRHINSAPMTVVLADADQPOLLPWIRPQPGD 344
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 83 LASVNVHGRTSMVEGIRVAQNIQKRT-VRHTNSCYFTWVAVDHKGKPREVPPLNDTEW 141
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 345 GERRYREAGARKKIRLDRKYIVSC 368
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 142 KRCFEAAHRRKYVLRQENHNPSG 165
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 5
US-09-134-001C-4696
; Sequence 4696, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4696
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4696

Query Match 5.9%; Score 186.5; DB 4; Length 177;
Best Local Similarity 30.6%; Pred. No. 1.3e-12;
Matches 56; Conservative 30; Mismatches 78; Indels 19; Gaps 4;

QY 17 FSNRTSRKSLRAGNDAMADGEGYRNFTEVOMQVLPCHTNOGELSVGQLKMTDT 76
DB 1 FMAINTNRKS-----KKMSDAKCHK-----TROVVPOTNNHHTFGGTLMANIDEI 46
QY 77 ACLSAERHAGCPCTASMDIYFEHTISVGQVNNKAVNRAFNSSMEVGIQVASEDLCS 136
DB 47 AAITAMKRAKPNVASTSDVDFLPITTTGDLSTEAANVSTAGISSMTCVQIVDDYK 106
QY 137 EKONVVCALATFYAREITK-VKLKQITPTEEEKMEHSA---AERRRRLVYADTI 191
DB 107 NERHMAISFLTFVALDSDGKPTSPDVEPEISVEKMHENAPQVARKERKESIDTI 166
QY 192 KDL 194
DB 167 EYL 169

RESULT 6
US-09-252-991A-31606
; Sequence 31606, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31606
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31606

Query Match 5.7%; Score 180.5; DB 4; Length 167;
Best Local Similarity 33.6%; Pred. No. 5.7e-12;
Matches 47; Conservative 25; Mismatches 67; Indels 1; Gaps 1;

QY 46 EVOMSQLVLPCHTNOGELSVGQLKMTDTTACLSAERHAGCPCTASMDIYFEHTISV 105
DB 8 QLTMSVLTMPDRANFGNVHGGALLKLDEVAFAACKRYAGRVVTLSDVYFRPDIHV 67

QY 106 GOVNNIKAVNRAFNSSMEVGIQVASEDLCSKQMNVCALATFYAREITK-VKLKQIT 164
DB 68 GELVTFIASVNYGRTIMEVGKVMYNTENHERYVRHINSCEFTVWAMDORPVAVPLE 127
QY 165 PRTEEXMEHSAERRMR 184
DB 128 PETAIEKRYAQAALAREQR 147

RESULT 7
US-09-540-236-2636
; Sequence 2636, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARI
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2636
; LENGTH: 162
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2636

Query Match 4.5%; Score 143; DB 4; Length 162;
Best Local Similarity 38.1%; Pred. No. 1e-07;
Matches 32; Conservative 17; Mismatches 35; Indels 0; Gaps 0;

QY 45 TEVOMSQLVLPCHTNOGELSVGQLKMTDTTACLSAERHAGCPCTASMDIYFEHTIS 104
DB 60 TDLNITVMTDMANFSGNAGGTLIKLDDVAAACASRYAGVYVTLSDVQVKEPLY 119
QY 105 VGQVNNIKAVNRAFNSSMEVGIQ 128
DB 120 VGEIVTFIARINHVQTTSMEVGIR 143

RESULT 8
US-09-543-681A-4296
; Sequence 4296, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4296
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4296

Query Match 4.2%; Score 135; DB 4; Length 143;
Best Local Similarity 31.2%; Pred. No. 6.7e-07;
Matches 39; Conservative 14; Mismatches 62; Indels 10; Gaps 3;

QY 218 TRVESVELVLPNHNQNTFGQIMAMNENTATIASRLCAHPTLKALIEHFHFGPSQ 277
DB 22 TRVSKV--LPTTNNHSTLFGETALAMWDSPLTATPCKRKLVTYSTKINFNHPLP 79
QY 278 VGBRLVTKAIVNNAFSGHMEVGVCVAYRQEAETHRRHINSAFMEFVLADAD-----D 330
DB 80 SGTIIILVEVIRVGRGTSILTVNVSIFLEMYVAGREVIHGQF--NVVAADHGKPTPLFD 138
QY 331 QPQLL 335

DB 139 KPSIL 143

RESULT 9
US-09-328-352-5165
Sequence 5165, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5165
LENGTH: 145
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5165

Query Match
Best Local Similarity 3.8%; Score 121; DB 4; Length 145;
Matches 35; Conservative 32; Mismatches 63; Indels 18; Gaps 3;

DB 39 EGRNPTEV-QMSQVLPCHTNORGEISVQQLKMDITTAQLSAERHAGPCVTASMDDI 97
DB 8 EATYKPEGLISTQIAMPADTNWSGVDFGGMIVSQMDLAGAIIHAFRSRGCATISINQM 67
QY 98 YFHTISVGVVNIKAKVNAFNSMEVGVQVASEDCSRKQNVCAALTFVA----- 151
DB 68 TELVPKVGFSVISCYTIKLVNGTSLQWGLEWSDSHDSRAPRIKTEGVTFVAVDYKGN 127
QY 152 RREITVKLQKQTPRTEEFMEHSVAAE 179
DB 128 KRQI-----PEAKQKFLISGE 144

RESULT 10
US-08-691-814B-6
Sequence 6, Application US/08691814B
Patent No. 5981218
GENERAL INFORMATION:
APPLICANT: Rio, Marie-Christine
APPLICANT: Tomasetto, Catherine
APPLICANT: Baaset, Paul
APPLICANT: Byrne, Jennifer
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Ave, NW, Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,814B
FILING DATE: 31-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/002,183
FILING DATE: 09-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 1383.0090001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2543
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 445 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-691-814B-6

Query Match
Best Local Similarity 3.7%; Score 118; DB 2; Length 445;
Matches 70; Conservative 53; Mismatches 97; Indels 130; Gaps 17;

QY 282 LVKAIYN-----NAFHSSEVGVCEVAYKQEAETHRRHNSAFMTFVILDADQPOLLP 336
DB 135 LIVKVIILSEILSKAGFYLLPIVSFVLAW-----LETWFLDFKVL----- 174
QY 337 WIRPQDGGRRY-----REASARKKI 358
DB 175 ----PQAEERRWTLAAVAVARGLPFGALSBGQFYSPBSFAGSDNESDEAVAGKSF 231
QY 359 RL-DRXYIVSCQTEVPLSVWDPNSQVYLSYNNVSLKMLVAKDNVWL--SEISQVRL 415
DB 232 SAQERXYIROQKEXTAVVD-----QILAQENKFKFENNEYGDT-V 271
QY 416 YTEDEKFLSFNHEMVHV- -AAQAFLLSDLRQPE-----NDKYSVELVQVDEDD 469
DB 272 YTLR---VPFHKFTFLKTFPLPCPAELVYQEVILQPERVWLNMKVTAQOIIQRVBDNT 327
QY 470 AI-YHTSPALGGHTKPODFVIL--ASRRPCONGDPY-----VIALESVTLPTRETPY 522
DB 328 LIIDVDSAGAGGVSPBDVNVRIERRR---DRYLSGATGTSARPPH-----Y 378
QY 523 RRGETLCSGFCILRBGDLTCCQVRSVLTSLVSAAGFSWGLESRSKR 572
DB 379 VRGENGRGRTIVLKSASNPVCTFV-----WILNTDLKGR 413

RESULT 11
US-09-252-991A-18251
Sequence 18251, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 18251
LENGTH: 312
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18251

Query Match
Best Local Similarity 3.6%; Score 113; DB 4; Length 312;
Matches 26; Conservative 8; Mismatches 36; Indels 8; Gaps 1;

QY 212 MYPEKTRVESVEL-----VLPRHANQGNTEFGQIMAMENYATIAASPLCAHPT 263
DB 83 MTPREGELIARTELSTRTKAVFPPTNNHNTLROGTALAMWDEVFTATRFCLPLV 142
QY 264 LKAIEMFHRGPSQVQDR 281

Db 143 TVSTDRIDFNHPICGAR 160

RESULT 12

US-09-543-681A-4796
; Sequence 4796, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4796
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4796

Query Match 3.5%; Score 112.5; DB 4; Length 146;

Best Local Similarity 26.5%; Pred. No. 0.00026;
Matches 36; Conservative 30; Mismatches 61; Indels 9; Gaps 2;

QY 46 EVQMSQVLPCHTNORGELISVGLIKWIDTTACLSAERHAGCPCTYASMDIYFEHTISV 105

Db 14 ELVRLTLPADTNANGDIFGGWIMSQMDIGAILAKEIALGRVTVSVTGTFLKPYAV 73

QY 106 GOVNNIRAKVNRANSSMEVGIOVASIEDLSE---KQNVCKALATFARREITVKLKQ 162

Db 74 GDVVCYARCLKTGSSITVINEVWKVATEPVQGRYRATBAYFTYVAVDENNSAR--- 130

QY 163 ITPREEKMEHSVAA 178

Db 131 ---RLPBGKTNFTLAS 143

RESULT 13

US-09-489-039A-7629
; Sequence 7629, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7629
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7629

Query Match 3.4%; Score 109.5; DB 4; Length 133;

Best Local Similarity 27.5%; Pred. No. 0.00048;
Matches 30; Conservative 26; Mismatches 50; Indels 3; Gaps 1;

QY 46 EVQMSQVLPCHTNORGELISVGLIKWIDTTACLSAERHAGCPCTYASMDIYFEHTISV 105

Db 12 EMVRLTLPADTNANGDIFGGWIMSQMDIGAILAKEIALGRVTVSVTGTFLKPYAV 71

QY 106 GOVNNIRAKVNRANSSMEVGIOVASIEDLSE---KQNVCKALATFVA 151

Db 72 GDVVCYARCLKTGSSITVINEVWKVATEPVQGRYRATBAYFTYVAV 120

RESULT 14

US-08-826-267-2
; Sequence 2, Application US/08826267
; Patent No. 5994070
; GENERAL INFORMATION:
; APPLICANT: Streuli, Michel
; TITLE OF INVENTION: NO. 5994070e1 TRIO Molecules and Uses Related Thereto
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHAYE & COCKFIELD
STREET: 28 Scate Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,267
FILING DATE: 1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,214
FILING DATE: 27 MARCH (1996)
ATTORNEY/AGENT INFORMATION:
NAME: Amy B. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2860 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-826-267-2

Query Match 3.3%; Score 106.5; DB 2; Length 2860;

Best Local Similarity 19.8%; Pred. No. 0.3;
Matches 126; Conservative 89; Mismatches 227; Indels 193; Gaps 32;

QY 22 SRKSLRAGNSAMADGEGYRNP-----TVQMSQVLPCHTNORGELISVGLIK 71

Db 1087 SAKQLEWTHD---NGSEFYLSHTSTGSSIOHTBELKEHEFEQITAKQTERYKLIQ 1142

QY 72 WIDTTACLSAERHAG--CPCTYASMDIYFEHTISVGVVNNIRAKVNRANSSMEVGIOV 129

Db 1143 LAD-GFCCKGHAHAIEIKCVTA-VDKRYRPSLSM-----EKYRSLKALGI 1189

QY 130 ASEDICSEKQNVCKALATFVARREITVKLKQITPRIEEKMEHSVAAERRMVLVYD 189

Db 1190 SSDSNKSSKSLQ---LDIIPASIPGSEVKLRDAHAELNEERKSA---RRKEPTIME 1240

QY 190 TIKDLNCAIQGDLSESDCSRMVPAEKTREVSVELVLPFPHNHQCNTEFG--QTMAMWE 247

Db 1241 LIQ---TEKAYVRD--RECDTYMENT--SGVEIEIPGLINKLILFGNNOIEIYEFIN 1293

QY 248 NVATIAASRLCPAHTPLXAIEMFHRGSPQVGDRLVLAIVNAPGHSYEVGC----- 301

Db 1294 NI-----FLKELEKYE-QLPEDVG---HCFVTWADKFCQYVTVYCKNKPDG 1334

QY 302 VEAYIOEA-----ETHRRH-INSAPMTFVVDADDOQPOLLPITRQPGGGERYREASAR 355

Db 1335 TQILIEHGSYDEDIQORHGLANSISYLI-----KPVORI 1370

QY 356 KRIPLDRKTIYVCKQ-----TEVPLSVPMDSNOYVLSY-----NNVSLKMLVAK 401

Db 1371 TYQVLLKELLTCCEBGKEIKDGLFVMLSVKRRANDAMHLSLBEFDNIESQGLIILQ 1430

QY 402 DN-----VLSEFI--SQVRLTYLLEDDKFLSTHEMNVHVD 435
DB 1431 ESFOVWDKPTLIRKGRREHLLEFENSIVFSKEVKDSSGRSKYLKSLFTSELGTEHVE 1490
QY 436 A-----AAFLILSLDLRQPEMDHYRVSVELVQVDEDDAI---YH-- 474
DB 1491 GDPCKFALWGRTPTSNDKIVLKASSIENKODMIKHIREVIOERTIHLKGLKEPIHPIK 1550
QY 475 TSPAL-----GGHTKQDFVILASRRK-----PCDNGDPYVIALRSVTL 513
DB 1551 TAPATROKGRBDEGLDQGGSSQPTITSIASRSQNTLSDSKSGGCELTVIHDTA 1610
QY 514 PTHRETPYRGFTL-----CSGFLMREGDQ 540
DB 1611 CNSNEL-TIRRGQTEVLELRPHDKPDCVLTETDR 1644

RESULT 15

US-09-543-681A-5467
; Sequence 5467, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5467
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5467

Query Match 3.3%; Score 106; DB 4; Length 722;
Best Local Similarly 18.7%; Pred. No. 0.027;

Matches 118; Conservative 83; Mismatches 171; Indels 260; Gaps 31;

QY 64 LSYGOLLKXIDITACLSAERHAGPCVTASMDIYFENHTISVGQVNVITAKVNAFNSM 123
DB 115 VTIG-LMKTYEGAAFIADK--SGAKYIPVRIEGAESYFSRVRKTLNITAKL----- 163
QY 124 EVGIQVASEDLCEKQNNVCKALATFVARREITKVKLKOTPRTEEXKMEHVAERRM 183
DB 164 -----FKITIK-VLPAVDLPMPTAKSAERRRL 191
QY 184 RLIVYADTIDLANCAIQGLDESRCRNVPAEKTRVESYELVLPphanQNTFGQIM 243
DB 192 S--GEKLRDIMDAIMQ----- 206
QY 244 AMWENVALTAASRLCAHPTLKALEMFRGSPQVGRVLKALVN-----NAPKSM 296
DB 207 -----ARPIITLYEAF-IQGMQFGRFSEPLISDIMKEDSYOGLKKML 249
QY 297 EVGVCEAYROEAETRRRHINSAFMTFVVL-DADQPOLPMIRPQGD-GERRYREASA 354
DB 250 GVSRIERTERTPERIRIGLLPNTTVVALFGATMRQVYAMINTYAGSIGVONAKMAAS 309
QY 355 RKKILDRKYIVSCKQTEVPLSVMDPSNQVYLsynNVSLKMLVAKDN--WV----- 405
DB 310 IKTIPTSRQFLKGNLHITPEQTP--DANWIIYL-----EDLKDSITREDKRWIAQHLTP 362
QY 406 -----LSSEISQVRLTYLED-----DKFLS-- 425
DB 363 HKQMAAQKSTDAVILFTSGSGCTPKGVVHSHSLANVEQIR--AIADFSPRDKEMAL 420
QY 426 --FH-MENVVHVA-----AAFLILSLDLRQPEMDHYRVS-ELVQVDEDDAIHYVTS 476
DB 421 PLFHAFGLTVSVLAPICLAGRVFLYPSPL-----HYRVPELY--YDQNCVTLFGTS 470

QY 477 PALGGHTK---PODFVILASRRKPCDNGDPYVIALRSVTLPTHRETPYRGFTLC--SGF 532
DB 471 TFLNGYGFAPHPYFARLR-----YVVA-----GAEKLSSESTR 503
QY 533 CLMR-----EGDOLTKCQWVSLTELVSASGFYSWGLESRSGRRSDGWNGLAGG 584
DB 504 VLMQDKFGIRILEGVGVTECAPV-VSINVPMASAKPY-----TVGRLLPGMBGRLL-- 551
QY 585 HLSTLKAIPVAKI-----NSRPGYIQ 605
DB 552 -----IPISGITDGRLOLRGPNIMGYLR 576

Search completed: August 12, 2004, 15:46:22
Job time : 20 secs

Blank Sheet

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 15:40:41 ; Search time 13 Seconds

(without alignments)
2431.276 Million cell updates/sec

Title: US-10-055-624B-2

Sequence: 1 MIGNVGNHRLRGASVPSNR.....TLKAIPIVAKINSRFGYLQDT 607

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3181	100.0	607	1	THEA_HUMAN
2	2530	79.5	594	1	THEA_MOUSE
3	1242.5	39.1	555	1	CACH_HUMAN
4	1240.5	39.0	556	1	CACH_MOUSE
5	1214.5	38.2	556	1	CACH_RAT
6	225	7.1	338	1	BACH_RAT
7	223.5	7.0	155	1	Y654_CHLNP
8	221	6.9	381	1	BACH_MOUSE
9	214.5	6.7	160	1	Y535_CHLTR
10	207.5	6.5	159	1	Y822_CHLMU
11	202	6.4	380	1	BACH_HUMAN
12	177	5.6	174	1	YKHA_BACST
13	175.5	5.5	179	1	YKHA_BACST
14	145	4.6	439	1	AC48_MOUSE
15	134.5	4.2	149	1	YCIA_ZYMO
16	131	4.1	406	1	CG16_HUMAN
17	126	4.0	239	1	ML64_MOUSE
18	121	3.8	446	1	ML64_MOUSE
19	121	3.8	1711	1	CHDI_MOUSE
20	120.5	3.8	233	1	STR6_MOUSE
21	118	3.7	445	1	ML64_HUMAN
22	114	3.6	154	1	YCIA_HAEIN
23	113.5	3.6	281	1	STAR_CHICK
24	113	3.6	294	1	ML64_SALFO
25	113	3.6	1709	1	CHDI_HUMAN
26	107.5	3.4	135	1	Y254_BUCBP
27	106.5	3.3	3038	1	TRIO_HUMAN
28	105.5	3.3	220	1	STR6_HUMAN
29	104.5	3.3	1663	1	CO3_HUMAN
30	104	3.3	132	1	YCIA_ECOLI
31	104	3.3	289	1	STAR_XENLA
32	103	3.2	133	1	YCIA_SALTY
33	102.5	3.2	985	1	CAPP_RALSO

ALIGNMENTS

RESULT 1	ID	THEA_HUMAN	STANDARD	PRT	607 AA.	
AC	08WX14	075187	096D11	09H883		P35418 taenia soli
DT	28-FEB-2003	(Rel. 41, Created)				P57362 buchnera ap
DT	28-FEB-2003	(Rel. 41, Last sequence update)				P33731 canis famli
DT	10-OCT-2003	(Rel. 42, Last annotation update)				P07694 homo sapien
DE	Brown fat inducible thioesterase (EC 3.1.2.-)	(BFTT)				Q98218 molluscum c
DE	associated thioesterase.					Q98218 molluscum c
GN	THEA OR BFTT OR KIAA0707.					P28366 bacillus su
OS	Homo sapiens (human)					P34431 caenorhabdi
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					P42398 buchnera ap
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					Q89541 bradyrhizob
OX	NCBI_TaxID=9606;					Q8ms1 drosophila
RP	SEQUENCE FROM N.A. (ISOFORM 1 AND 2).					Q9qgn8 rattus norv
RX	MEDLINE=21552902; PubMed=11686000;					
RA	Adams S.H., Chui C., Schilbach S.L., Yu X.X., Goddard A.D.,					
RT	Grimaldi J.C., Lee J., Dowd P., Colman S., Lewin D.A.;					
RT	"BFTT, a unique acyl-CoA thioesterase induced in thermogenic brown					
RT	adipose tissue: cloning, organization of the human gene and					
RT	assessment of a potential link to obesity.";					
RL	Biochem. J. 360:135-142(2001).					
RL	[2]					
RP	SEQUENCE FROM N.A. (ISOFORM 1).					
RX	MEDLINE=98403880; PubMed=9734811;					
RA	Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,					
RT	Kotani H., Nomura N., Ohara O.;					
RT	"Prediction of the coding sequences of unidentified human genes. X.					
RT	The complete sequences of 100 new cDNA clones from brain which can					
RT	code for large proteins in vitro.";					
RL	DNA Res. 5:169-176(1998).					
RL	[3]					
RP	SEQUENCE FROM N.A. (ISOFORM 2).					
RX	MEDLINE=22388257; PubMed=12477932;					
RA	Klausner R.D., Collins F.S., Wagner L.H., Derge J.G.,					
RT	Shenmen C.M., Schuler G.D.,					
RT	Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.;					
RT	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,					
RT	Stachenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,					
RT	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,					
RT	Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,					
RT	Bosak S.A., McEwan P.J., McKernan K.D., Malek J.A., Gunaratne P.H.,					

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scheraga A., Schein J.B., Jones S.J.M., Maira M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: Has acyl-CoA thioesterase activity towards medium (C12)
 CC and long-chain (C18) fatty acyl-CoA substrates.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=BF11;
 CC IsoId=08MX14-1; Sequence=Displayed;
 CC Name=2; Synonyms=BF12;
 CC IsoId=08MX14-2; Sequence=VSP 000160;
 CC -1- TISSUE SPECIFICITY: Isoform 1 is predominantly expressed in
 CC skeletal muscle, liver, testis, stomach, spleen, lung and brain.
 CC Isoform 2 is predominantly expressed in kidney, uterus, hibernoma
 CC and white adipose tissue.
 CC -1- INDUCTION: By cold exposure and repressed by heat exposure.
 CC -1- SIMILARITY: Contains 2 acyl coenzyme A hydrolase domains.
 CC -1- SIMILARITY: Contains 1 START domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF416921; AAL40937.1; -;
 DR EMBL; AF416922; AAL40938.1; -;
 DR EMBL; AB014607; BAA31682.1; AUT_INIT.
 DR EMBL; AK023937; BAA14734.1; -;
 DR EMBL; BC001517; AAO1517.1; -;
 DR GeneW; HGNC:18156; THEA.
 DR MIM; 606803; -;
 DR GO; GO:0005737; C:cytoplasm; IC.
 DR GO; GO:0016291; F:acyl-CoA thioesterase activity; TAS.
 DR GO; GO:0006631; P:fatty acid metabolism; NAS.
 DR GO; GO:0007242; P:intracellular signaling cascade; NAS.
 DR GO; GO:0009266; P:response to temperature; ISS.
 DR InterPro; IPR002913; START.
 DR InterPro; IPR006683; Thioestr_suff.
 DR Pfam; PF03061; 4HBT; 2.
 DR Pfam; PF01852; START; 1.
 DR SMART; SMO0234; START; 1.
 DR PROSITE; PS00848; START; 1.
 KM Hydrolyase; Serine esterase; Repeat: Alternative splicing.
 FT DOMAIN 29 164 ACTYL-COA HYDROLASE 1.
 FT DOMAIN 205 336 ACTYL-COA HYDROLASE 2.
 FT DOMAIN 375 585 START.
 FT VARSPLIC 544 607
 FT CONFLICT 255 255 \$-> R (IN REF. 4).
 FT SEQUENCE 607 AA; 68492 MW; 12F2BCBA8A318C CRC64;
 Query Match 100.0%; Score 3181; DB 1; Length 607;
 Best Local Similarity 100.0%; Pred. No. 6.2e-244;
 Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 RGEISVQGLKMTIDTTACTLSAERHAGCPVTASMDITYFEHTISVGVNMIKAKVNAFN 120
 DB 61 RGEISVQGLKMTIDTTACTLSAERHAGCPVTASMDITYFEHTISVGVNMIKAKVNAFN 120
 QY 121 SSMVEGIVASSEDICSEKQNVNVCALATFVARREITVKLKQITPRTEEMESVAER 180
 DB 121 SSMVEGIVASSEDICSEKQNVNVCALATFVARREITVKLKQITPRTEEMESVAER 180
 QY 181 RRMRLVYADITKOLLNCAIQGDIESRDCEKMPAEKTRVESVLPVPHANHGNTFGG 240
 DB 181 RRMRLVYADITKOLLNCAIQGDIESRDCEKMPAEKTRVESVLPVPHANHGNTFGG 240
 QY 241 QIYAMMENVATIAASRLCRAHPTLKAIEMFHRPSQVDELVKAIYNNAFKSMVEGV 300
 DB 241 QIYAMMENVATIAASRLCRAHPTLKAIEMFHRPSQVDELVKAIYNNAFKSMVEGV 300
 QY 301 CVEAYRGEAETHRRHINSAPMTFVLDADDPQLLPWIRPQDGERRYRASARKIRL 360
 DB 301 CVEAYRGEAETHRRHINSAPMTFVLDADDPQLLPWIRPQDGERRYRASARKIRL 360
 QY 361 DRKTYVCKQTEVPLSVPMPSNQVYLSYNNVSLKMLVAKDNWVLSSEISOVRLTYLED 420
 DB 361 DRKTYVCKQTEVPLSVPMPSNQVYLSYNNVSLKMLVAKDNWVLSSEISOVRLTYLED 420
 QY 421 DKFLSFHMEVHVHDAQAFLILSDLRQRPMDKHYSVELVQVDEDDAIYHVTSPALG 480
 DB 421 DKFLSFHMEVHVHDAQAFLILSDLRQRPMDKHYSVELVQVDEDDAIYHVTSPALG 480
 QY 481 GHTXPQDFVLIASRRKPCDNDPYVIALRSVTLPTHRETPYRGETLCSGFLMRBEDQ 540
 DB 481 GHTXPQDFVLIASRRKPCDNDPYVIALRSVTLPTHRETPYRGETLCSGFLMRBEDQ 540
 QY 541 LTKCCWRVSLTELVSAGFYSGWESKGRSGNMGKLAGHLSLTKAIPVAKINSR 600
 DB 541 LTKCCWRVSLTELVSAGFYSGWESKGRSGNMGKLAGHLSLTKAIPVAKINSR 600
 QY 601 FGYLQDT 607
 DB 601 FGYLQDT 607
 RESULT 2
 THEA_MOUSE
 ID THEA_MOUSE STANDARD; PRT; 594 AA.
 AC Q8VHQ9; 2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Brown fat inducible thioesterase (EC 3.1.2.-) (BFIT) (Adipose
 DE associated thioesterase).
 GN THEA OR BFIT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RX MEDLINE=21552902; PubMed=1169600;
 RA Adams S.H., Chui C., Schilbach S.U., Yu X.X., Goddard A.D.,
 RA Grimaldi J.C., Lee J., Dowd P., Colman S., Lewin D.A.,
 RT "BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown
 RT adipose tissue: cloning, organization of the human gene and
 RT assessment of a potential link to obesity.";
 RT Biochem. J. 360:135-142(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; Tissue=Kidney;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

FT	DOMAIN	327	536	START
5Q	SEQUENCE	556 AA;	61761 MW;	9FF9C487BBCB812 CRC64;
	Query Match	39.0%;	Score 1240.5;	DB 1; Length 556;
	Best Local Similarity	50.8%;	Pred. No. 2.1e-90;	
	Matches 253;	Conservative 93;	Mismatches 139;	Indels 13; Gaps 7;
QY	44	PTENVQSGVLVPCHTNORGELSVGGLKMTDTTACLSAEHRNAGCCVYASMDIYEERTI	103	
DB	7	PGEVLMSQAIOPAHADSRGELSAQGLKMTDTTACLSAEHRNAGCCVYASMDIIFEEDTA	66	
QY	104	SVGVQVNVKAKAVNRFAINSMEVGIQVASEDCSEKQMVVCALATFVAR---REITTKL	160	
DB	67	RIGQITTRAKYTRAFSTSMELISKVIYQDFGTGIGLCLVAFFSVKAPVKE--KVHL	124	
QY	161	KQITPREEKMEHSVAERRRMRVLVADTKDLIANCAIQDLESRDCSRMPVPAEKTRV	220	
DB	125	KPVLLQTEQGEVHNHMASERRKVRLOHNTNTNNIMKESRRSDSI CNEEBGATMTGTSV	184	
QY	221	ESVELVLPPEAHNQGNTGGQIMAMENVATTASRLCRAHPHTLKALIMFHRGSGYQD	280	
DB	185	QSEIVLPPPHANNHGNITFGQIMAMENVATTASRLCHGPFLKSVDMFKFRGPGSTVG	244	
QY	281	RLVVKATVNNAPFKSMSEVGCVEAY--ROKETHRRHNTNSAPMFVVLVADADQPOL--P	336	
DB	245	RLVNSALVNNTPFQSVSEVGVAREADCCQEMAGCGRHNSAPLYNAA--DDQELITFP	302	
QY	337	WIRPQDGERRERBASARKKIRLDRKTVSCQTEVEVLSPWDPNSOVLYSNVNSSLK	396	
DB	303	RIGPISKDPRRYGCAIARRRIRLGRKKVYIHHK--EVLPSAQMDISKSGISLNTVEMLK	361	
QY	397	MLVAKDNVLSSEISQVRLTYLEDDKFLSPFHEMVHVDAQAQILLSDLRQRPBWDKHY	456	
DB	362	NLASGSGWEITTTLEKIKITYLEBODAISVKEKLVGSPAHIAVHLSDLTRPLMDPHY	421	
QY	457	RSVELVQGVDDDDAIVHTYSALGHTPPOPFVILASRKXCNDGDPVILALRSTLPTA	516	
DB	422	ISCEVITDVSDDDQIYITGCVVNG-DKPKDFVLVSRKPLKNNITYALRSVLPSPV	480	
QY	517	RETPYERGETLCSGFLC 534		
DB	481	PSFPQIRSEVYCAGFLI 498		
RESULT 5				
CACH_RAT	ID	CACH_RAT	STANDARD;	PRT; 556 AA.
AC	Q99NB7;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, last sequence update)			
DT	10-OCT-2003 (Rel. 42, last annotation update)			
DE	Cyclophilin acetyl-CoA hydrolase 1 (EC 3.1.2.1) (CACH-1) (RACH).			
GN	CACPHI OR CACH.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.S.A. AND SEQUENCE OF 161-174 AND 352-364.			
RC	STRAIN=Donryu; TISSUE=Liver;			
RX	MEDLINE=2123026; PubMed=11322891;			
RA	Sumatrasu N., Okamoto K., Shibata K., Nakanishi Y., Isohashi F.;			
RT	"Molecular cloning and functional expression of rat liver cytosolic			
RT	acetyl-CoA hydrolase.";			
RL	Eur. J. Biochem. 268:2700-2709 (2001).			
CC	-1- FUNCTION: Hydrolyzes acetyl-CoA to acetate and CoA.			
CC	-1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O = CoA + acetate.			
CC	-1- ENZYME REGULATION: Allosterically regulated by ATP (activator) and			
CC	ADP (inhibitor).			
CC	-1- PATHWAY: Probably involved in fatty acid metabolism.			
CC	-1- SUBUNIT: Active homodimer or homotetramer at room temperature and			
CC	inactive monomer at low temperature.			
CC	-1- SUBCELLULAR LOCATION: Cyttoplasmic.			

ID	BACH RAT	STANDARD	PRT	338 AA
AC	Q64553	008652; 003041		
DT	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, last sequence update)		
DT	15-MAR-2004	(Rel. 43, last annotation update)		


```

RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
RT pneumoniae A39."
RL Nucleic Acids Res. 28:1397-1406 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CML029 from USA."
RL Nucleic Acids Res. 28:2311-2314 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=TW-183;
RA Gang M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
RT other Chlamydia strains based on whole genome sequence analysis."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL: AB001648; AAD18793.1; -
DR EMBL: AB002172; AAF37977.1; -
DR EMBL: AF002547; BAA98861.1; -
DR EMBL: AB017159; AAP98609.1; -
DR PIR: C86572; C86572.
DR PIR: G72051; G72051.
DR TIGR: CP0093; -
DR InterPro: IPR006683; Thioestr_supf.
DR Pfam: PF03061; 4HBT; 1.
KW Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 155 AA; 17883 MW; FE1483EA30323AA CRC64;

Query Match 7.0%; Score 223.5; DB 1; Length 155;
Best Local Similarity 35.5%; Pred. No. 1.1e-10;
Matches 50; Conservative 29; Mismatches 55; Indels 7; Gaps 2;

QY 226 VLPPHANGNTFGGOIMAMENAVTIAASRLGRAHPTLKAIMEFHRGSGVGDRLYIK 285
DB 19 IFPNDIANANTVGGILMSLDRLALVVAKRHTESCVTAFPDALRYPAYMGENTLCK 78
QY 286 AIVNNAFKISMVEGVCEVA---YRQEAETRRRHINSAMFTFVVLADDDQPOLLPWIRPOP 342
DB 79 AAVNRTRWRTSLFEGVVKWMAENIKQE---RHHTSAYFFVAVNEDNGIPVHQIVPT 134
QY 343 GDSGRYRERSAKKTRIDRK 363
DB 135 PEKKRRYNEADRRRRQRLLEIK 155

RESULT 8
BACH_MOUSE
ID BACH_MOUSE STANDARD; PRT; 381 AA.
AC 091V12;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytosolic acyl coenzyme A thioester hydrolase (EC 3.1.2.2) (long chain
DE acyl-CoA thioester hydrolase) (CTR-II) [Brain acyl-CoA hydrolase].
OS BACH.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

```

```

RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC STRAIN=ICR; TISSUE=Brain;
RA Takagi-Sakuma M., Kuramochi Y., Kitahara M., Emori R., Kuroda J.,
RA Hiratsuka K., Nagae Y., Watanabe T., Suga T., Unji Y.;
RT "Mouse brain long-chain acyl-CoA hydrolase."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS B AND C).
RC STRAIN=ICR; TISSUE=Brain;
RA Yamada J., Kuramochi Y.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Colon;
RX MEDLINE=2338257; PubMed=12477932;
RA Strausberg R.L., Peingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heile F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Kana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Feibey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smillius D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- FUNCTION: May play an important physiological function in brain.
CC May play a regulatory role by modulating the cellular levels of
CC fatty acyl-CoA ligands for certain transcription factors as well
CC as the substrates for fatty acid metabolizing enzymes,
CC contributing to lipid homeostasis. Has broad specificity, active
CC towards fatty acyl-CoAs with chain-lengths of C8-C18. Has a
CC maximal activity toward palmitoyl-CoA.
CC -1- CATALYTIC ACTIVITY: Palmitoyl-CoA + H(2)O = CoA + palmitate.
CC -1- SUBUNIT: Homodimer (Probable).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=B; Synonyms=mbACHb;
CC IsoId=Q91V12-1; Sequence=Displayed;
CC Name=A; Synonyms=mbACHa;
CC IsoId=Q91V12-2; Sequence=VSP_000158;
CC Name=C; Synonyms=mbACHc;
CC IsoId=Q91V12-3; Sequence=VSP_000157;
CC -1- SIMILARITY: Contains 2 acyl coenzyme A hydrolase domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL: AB049821; BAB61731.1; -
DR EMBL: AB088411; BAC20217.1; -
DR EMBL: AB088412; BAC20218.1; -
DR EMBL: BC013507; AAI13507.1; -
DR MGD: MGI:1917275; Bach.
DR GO: GO:0005829; C:cytosol; IDA.
DR GO: GO:0009062; P:fatty acid catabolism; IDA.
DR InterPro: IPR008994; Nucleic_acid_OB.
DR InterPro: IPR006683; Thioestr_supf.

```


DR Pfam; PF03061; 4HBT; 2
 KW Hydrolyase; Serine esterase; Repeat; Alternative splicing.
 FT DOMAIN 44 178 ACYL-COA HYDROLASE 1.
 FT DOMAIN 210 346 ACYL-COA HYDROLASE 2.
 FT VARSPPLIC 1 59 MKLVGTLRLEVEGRQVAFSLTPGQSGRLKTFMAARA
 VRTRADHOKLGHCVTMGR -> MTLTRHALRVLKREKVE
 AYLRKVKO (in isoform C).
 /FTID=VSP_000157.
 FT VARSPPLIC 1 58 MKLVGTLRLEVEGRQVAFSLTPGQSGRLKTFMAARA
 VRTRADHOKLGHCVTMG -> MSGPTDTPAIGIC (in
 isoform A).
 /FTID=VSP_000158.
 FT SEQUENCE 381 AA; 42537 MW; 813852DBB6834C4 CRC64;
 Query Match 6.9%; Score 221; DB 1; Length 381;
 Best Local Similarity 22.6%; Pred. No. 6e-10;
 Matches 74; Conservative 64; Mismatches 154; Indels 36; Gaps 7;
 QY 47 VQMSQVLPHCHTNGRGLSVGLKMTDTACLSAERAG-----CPCVTASMDTYFE 100
 DB 55 VTMGRIMPDADANVAGVHGSTILKMLEAGAITSTHCSNGRRCVVALARVERIDFL 114
 QY 101 HTLSVGVVNIKAVNFAFNSMVEGIVASEDLCEKQNVCKALATFV--ARRRITKV 158
 DB 115 SPMGIGFAVHASAETITTSKSHVEQVAVMSENILTGKTLTKATLMYVPLSKVDKV 174
 QY 159 ----KMKQITRTTEKKMESHVAAERRRMELV--ADITKDLNCAIQGLSDGRDSRM 212
 DB 175 LEVPPIVYLROEHEEGKRKRYAOKLEBETKMRNGDVPDLN----- 218
 QY 213 VPARKT-----RVESEVELVPPHANHOGNTFGQIMAMENVAITIASRLCRAHPTLKAE 268
 DB 219 -PEPVNTSYSSSLIHVPSDCTLHGVRHGVTKMLMDRYAGVIAHCHKNITVTSVD 277
 QY 269 MEHFRGSPQVDRVLVLAIVNAFHSMEVGVCEAYEQEAEHRRH-INSAPMTFVLVD 327
 DB 278 AINPHDKIRKGCVITISGRMTFTSKNSWEIEVLVDADPVNDSQKRYAASAFYVSLN 337
 QY 328 ADDQQLLPMTIRPQDGERRYRARSAR 355
 DB 338 QEGKMPVPQVLVPEDEKRRFEKGR 365
 RESULT 9
 Y535 CHLUTR STANDARD; PRT; 160 AA.
 ID Y535 CHLUTR
 AC 084540;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative acyl-CoA thioester hydrolase CT535 (EC 3.1.2.-).
 OS Chlamydia trachomatis.
 NC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 NCBI_TaxID=813;
 RX STRAIN=D/UM-3/Cx;
 MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kallman S., Lamell C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis";
 RL Science 282:754-759(1998).
 CC -I- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC EMBL; AE001324; AAC68137.1; -.
 DR PIR; F71502; F71502. Thioestr_suff.
 DR InterPro; IPR006683; Thioestr_suff.
 DR Pfam; PF03061; 4HBT; 1.
 KW Hypothetical protein; Hydrolase; Complete proteome.
 SQ SEQUENCE 160 AA; 18526 MW; 90881F56F13BA6F2 CRC64;
 Query Match 6.7%; Score 214.5; DB 1; Length 160;
 Best Local Similarity 34.1%; Pred. No. 5.8e-10;
 Matches 47; Conservative 30; Mismatches 54; Indels 7; Gaps 2;
 QY 226 VLPPHANHOGNTFGQIMAMENVAITIASRLCRAHPTLKAEHFRGSPQVGDRLVLK 285
 DB 25 IFPVDLNTNNTIFGLMSLIDRLALVVAEHCSICTATAVDAMRFAPAYMGENILIC 84
 QY 286 AIVNAFHSMEVGVCEA--YRQEAETHRRHINSAPMTFVLADQQLPMTIRPQ 342
 DB 85 ASVNSRWSRTSLSEGVKVAENITKOE--HRHITSATFTFVAVDKNNSPVEVDELIPS 140
 QY 343 GDGERRYRARSARKIRL 360
 DB 141 QEEIRRRERADQRRALRL 158
 RESULT 10
 Y822 CHLMU STANDARD; PRT; 159 AA.
 ID Y822 CHLMU
 AC Q9PJX7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative acyl-CoA thioester hydrolase TC0822 (EC 3.1.2.-).
 OS Chlamydia muridarum.
 NC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 NCBI_TaxID=83560;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uteerback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClary G., Salzberg S.L.,
 RA Eisen J., Fraser C.W.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -I- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC EMBL; AE002347; AAF39624.1; -.
 DR PIR; C81662; C81662.
 DR TIGR; TC0822; -.
 DR InterPro; IPR006683; Thioestr_suff.
 DR Pfam; PF03061; 4HBT; 1.
 KW Hypothetical protein; Hydrolase; Complete proteome.
 SQ SEQUENCE 159 AA; 18248 MW; EFNAD7BDCB692B11 CRC64;
 Query Match 6.5%; Score 207.5; DB 1; Length 159;
 Best Local Similarity 33.3%; Pred. No. 2.1e-09;
 Matches 46; Conservative 30; Mismatches 55; Indels 7; Gaps 2;
 QY 226 VLPPHANHOGNTFGQIMAMENVAITIASRLCRAHPTLKAEHFRGSPQVGDRLVLK 285

```

Db      24  I PPDNDNTNNTTITFGGLIMSLDLRLAIWAERFCESICVTALVDAAVFAPYAMGENTLICC 83
Oy      286 AIYNNAFKHSMENGVCEVA---YRGAEIHRRHINSAMFTFVLDDADQPOLWIRPOP 342
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      84  ASVNRSSRRTSLEGVAVMAENIKOE---QRHTSAFFTFVADENNSPVEIPELVPET 139
Oy      343 GDGGRRRYEASARKIRL 360
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      140 QEFIRRFHEADQRRASRL 157

RESULT 11
BACH HUMAN
ID BACH HUMAN STANDARD; PRT: 380 AA.
AC 000154; 043703; Q9UJMG; Q9Y539; Q9Y544;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytosolic acyl coenzyme A thioester hydrolase (EC 3.1.2.2) (Long chain
acyl-CoA thioester hydrolase) (CTE-II) (Brain acyl-CoA hydrolase).
GN BACH.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN RP SEQUENCE FROM N.A. (ISOFORM 4), AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=20047053; PubMed=10578051;
RA Yamada U., Kurita A., Hirata M., Taniguchi T., Takama H., Furinata T.,
RA Shiratori K., Iida N., Takagi-Sakuma M., Watanabe T., Kurosaki K.,
RA Endo T., Suga T.;
RT "Purification, molecular cloning, and genomic organization of human
RL brain long-chain acyl-CoA hydrolase."
RJ J Biochem. 126:1013-1019(1999).
[2]
RN RP SEQUENCE FROM N.A.
RC TISSUE=Hippocampus;
RA Hajira A.K., Uhler M.D., Larkins L.K.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 5 AND 6).
RA Yamada U., Kuramochi Y.;
RL "Human brain acyl-CoA hydrolase isoforms."
RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Staussberg R.D., Collins F.S., Wagner L., Shemen C.M., Schutler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stepienon M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosnak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kelleman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butefisch J.Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Scherach A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences ";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[5]
RN RP SEQUENCE OF 1-247 FROM N.A.
RA Howden P.;

```

```

RL Submitted (JUN-1999) to the EMBL/Genbank/DBJP databases.
CC -I- FUNCTION: May play an important physiological function in brain.
      May play a regulatory role by modulating the cellular levels of
      fatty acyl-CoA ligands for certain transcription factors as well
      as the substrates for fatty acid metabolizing enzymes,
      contributing to lipid homeostasis. Has broad specificity, active
      towards fatty acyl-CoAs with chain-lengths of C8-C18. Has a
      maximal activity toward palmitoyl-CoA.
CC -I- CATALYTIC ACTIVITY: Palmitoyl-CoA + H(2)O = CoA + palmitate.
CC -I- SUBUNIT: Homodimer (Probable).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- ALTERNATIVE PRODUCTS:
      Event=Alternative splicing; Named isoforms=6;
      Name=1; Synonyms=B, hBACHD;
      IsoId=000154-1; Sequence=D displayed;
      Name=2; Synonyms=A-X, hBACHA-X;
      IsoId=000154-2; Sequence=VSP_000152, VSP_000155, VSP_000156;
      Name=3; Synonyms=A-X1, hBACHA-X1;
      IsoId=000154-3; Sequence=VSP_000154;
      Name=4; Synonyms=A, hBACHA;
      IsoId=000154-4; Sequence=VSP_000152;
      Name=5; Synonyms=C, hBACHC;
      IsoId=000154-5; Sequence=VSP_000151;
      Name=6; Synonyms=D, hBACHd;
      IsoId=000154-6; Sequence=VSP_000153;
CC -I- SIMILARITY: Contains 2 acyl coenzyme A hydrolase domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D88894; BA824350.1; -.
DR EMBL; U91316; AB861211.1; -.
DR EMBL; AB074415; BAC20174.1; -.
DR EMBL; AB074417; BAC20175.1; -.
DR EMBL; AB074417; BAC20176.1; -.
DR EMBL; AB074418; BAC20177.1; -.
DR EMBL; AB074419; BAC20178.1; -.
DR EMBL; BC017365; AAH17365.1; -.
DR EMBL; AL031848; CAB46201.1; -.
DR EMBL; AL031848; CAB46202.1; -.
DR EMBL; AL031848; CAB46203.1; -.
DR PIR; JC7161; JC7161.
DR GO; GO:0005737; C:cytosolam; TAS.
DR GO; GO:0000062; F:acyl-CoA binding; TAS.
DR GO; GO:0003824; F:catalytic activity; TAS.
DR GO; GO:0006629; F:lipid metabolism; TAS.
DR InterPro; IPR006683; Thioestr_sufp.
DR Pfam; PF03061; 4HT; 2.
KW Hydrolyase; Serine esterase; Repeat; Alternative splicing.
FT DOMAIN 43 177 ACYL-COA HYDROLASE 1.
FT     209 345 MKLRLARLRICEFGQASSRRLVAGQGCVGRRGCCAPGV
FT     VARSLIC 1 58 VGPADLPKPGCACTIGR -> MLTLRLSLINLVKEVDRA
FT                                     CFGEKAKQ (in isoform 5).
FT /ftid=vsp_000151.
FT     VARSLIC 1 57 MKLLARLARLTCEPGRQASRRLVAGQGCVGRRGCCAPGV
FT                                     VGPADLPKPGCACTIG -> MSCPDVEPSALIOC (in
FT                                     isoform 2, isoform 3 and isoform 4).
FT /ftid=vsp_000152.
FT     VARSLIC 1 57 MKLLARLARLTCEPGRQASRRLVAGQGCVGRRGCCAPGV
FT                                     VGPADLPKPGCACTIG -> MAFOLS (in isoform
FT                                     6).
FT /ftid=vsp_000153.
FT     VARSLIC 287 380 GCVTITSGRMPTSTKSMEIVLVADDPVDSOKRYRAS
FT                                     APFTYVSLSOGSRSLPYQLVPETDEKKREPRGGAYLN
FT     KAKRGHAEPOR -> AHVMPPGAQHNTLAPESSSPTGTCKSL
FT     LRHHHLGTHDHLEQ (in isoform 3).
FT /ftid=vsp_000154.
```

```

FT VARSPLIC 287 288 GC -> AP (in isoform 2).
FT VARSPLIC 289 380 Missing (in isoform 2).
FT CONFLICT 371 380 /FTID=VSP_000156.
FT SEQUENCE 380 AA; 41796 MW; BDD75D62A60095EC CRC64;
SQ
Query Match
Best Local Similarity 21.9%; Pred. No. 1.9e-08;
Matches 71; Conservative 61; Mismatches 156; Indels 36; Gaps 6;

QY 51 QLVLPCHTNGRGLSVGQLTKMIDTTACTLSAERHAG-----CPVTASMDIYFHTIS 104
DB 58 RIMPPDDANVAGNVHGTILMKIEBAGAISTRHNSNGRVCVAALARVERTDLSMC 117
QY 105 VGVVNTKAKYKRNAPNSMEVGIQVASEDLCEKQMNCKALAFVAREITKYKQIT 164
DB 118 IGEVAHVSAEITLYSKHSVEVQVNVMSHNLTKAKTLWVPLSLKRVKLEVP 177
QY 165 P-----RTEEEKHESVAERBRRLVY--ADTKDLNCAIIGDLESRCRMVPAE 216
DB 178 PVTYSRQGEERGRYRQAQKLERETKMRNGDIVQVLA-----PEP 220
QY 217 KT----RVESVELVLPFHNHQNTFGQIMAMNENATIASRLCRAPTLKAIEMFHF 272
DB 221 NTVSYSQSLIHIVSPDCTLHGPHGVTMKMDVAGIYAARCKNITVATASVDALNF 280
QY 273 RGPQVQDRLVYLKATVNNAPHSMEVGVCAVNRQEAETRRH--INSAFMTFVIDADQ 331
DB 281 HDKIRKRGCVITISQKMTFTSNKSMIEVLVDADPVDSSQKRYAASAFYTVLSQGR 340
QY 332 POLPMTIRPQDGRRRYREASAR 355
DB 341 SLVPQLVPTEDRKKRPEBGKGR 364

RESULT 12
VLDL_HELPY
ID_VLDL_HELPY STANDARD; PRT; 174 AA.
AC 005729;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein vldl.
GN VLDL OR HP0891 OR JHP0824.
OS Helicobacter pylori (Campylobacter pylori), and
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210, 85963;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49503 / 60190;
RX MEDLINE=97284485; PubMed=9139899;
RA Cao P., Cover T.L.;
RT "High-level genetic diversity in the vldl chromosomal region of
RT Helicobacter pylori."
RL J. Bacteriol. 179:2852-2856 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter

```

```

RT pylori."
RL Nature 388:539-547 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J99;
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Tria-Nickelsen M., Mills D.K., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180 (1999).
CC -1- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U94318; AAC5242.1; -.
DR EMBL; AE000599; AAD07940.1; -.
DR EMBL; AE001512; AAD06410.1; -.
DR PIR; C64631; C64631.
DR PIR; T09451; T09451.
DR TIGR; HP0891; -.
DR InterPro; IPR006683; Thioestr_supf.
DR Pfam; PF03061; 4HBT; 1.
KT Hydrolase; Complete proteome.
FT CONFLICT 8
SQ SEQUENCE 174 AA; 19491 MW; 48D2C71C9AD2A534 CRC64;

Query Match
Best Local Similarity 31.4%; Pred. No. 6e-07;
Matches 44; Conservative 28; Mismatches 69; Indels 0; Gaps 0;

QY 45 TEVOMSQLVDPCHTNGRGLSVGQLTKMIDTTACTLSAERHAGCPVTASMDIYFHTIS 104
DB 22 TKLMSYLVVPTTANFNVMHGGELNLDKVAVCSYKCAKGVTLSDGVTEKYP 81
QY 105 VGVVNTKAKYKRNAPNSMEVGIQVASEDLCEKQMNCKALAFVAREITKYKQIT 164
DB 82 VGNLLTFLASINIVGNSTCEVGIKVSBDIKTRITHNNSCYFTWAVANGKPTMPKYE 141
QY 165 PRTEEEKHESVAERBRMR 184
DB 142 PTEVEIRRYEGALKRKEMR 161

RESULT 13
YKHA_BACSU
ID_YKHA_BACSU STANDARD; PRT; 179 AA.
AC P49851;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative acyl-CoA thioester hydrolase ykha (EC 3.1.2.-).
GN YKHA OR BSU1030.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96272254; PubMed=9682784;
RA Lacelle M., Kumano M., Kurita K., Yamane K., Zuber P., Nakano M.M.;
RT "Oxygen-controlled regulation of the flavohemoglobin gene in Bacillus
RT subtilis."
RL J. Bacteriol. 178:3803-3808 (1996).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Devine K.M.;
RT "Sequence of the Bacillus subtilis genome between xlyA and ykor.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Boris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.C., Comerton I.F., Cummings N.J., Daniel R.A.,
Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Funa S., Gallizzi A., Galleron N.,
Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaeer-Blanchard M., Klein C.,
Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Portolillo S., Prescott A.M.,
Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rooha E., Roche B., Rose M., Sadie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
Sekituchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takenchi M., Tamakoshi A., Tanaka T., Terpstra P., Togomori K.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viat P., Mambrot R., Wedler E., Wedler H., Weitzmeger T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zundin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:49-256(1997).
CC -1- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D78189; BAA11257.1; -
DR EMBL: AJ002571; CAA05583.1; ALT INIT.
DR EMBL: Z99110; CAB13160.1; ALT INIT.
DR Subtilast; BG11417; ykha.
DR InterPro; IPR006683; Thioestr_supf.
DR Pfam; PF03061; 4HBT; 1.
KW Hydrophobic protein; Hydrolase; Complete proteome.
SQ SEQUENCE 179 AA; 20065 MW; 8CB508A96A15B CRC64;

```

```

Db 132 VPETEEMHNTAVGRANER 152
|||||
RESULT 14
AC48 MOUSE STANDARD; PRT; 439 AA.
ID AC48_MOUSE
AC Q9R0X4; Q9WTJ0; Q9WU28;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 48 kDa acyl-CoA thioester hydrolase, mitochondrial precursor
DE (EC 3.1.2.-) (p48) (Mt-ACT48) (Protein U8).
GN ACAT2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=99315862; PubMed=10383425;
RA Poupon V., Begue B., Gagnon J., Dautry-Varsat A., Cerf-Bennussan N.,
RA Benmerrah A.;
RT "Molecular cloning and characterization of Mt-ACT48, a novel
RT mitochondrial acyl-CoA thioesterase.";
RL J. Biol. Chem. 274:19188-19194(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AKR.
RA Ishizuka Y., Mochizuki R., Tondoh N.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Active on long chain acyl-CoAs.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ238893; CAB45192.1; -
DR EMBL: AJ238894; CAB45255.1; -
DR EMBL: AB028898; BAA79193.1; -
DR MGD; MGI:1928939; Acate2.
DR InterPro; IPR006683; Thioestr_supf.
DR Pfam; PF03061; 4HBT; 1.
KW Hydrolase; Serine esterase; Mitochondrion; Transit peptide.
RN TRANSIT 1 21
MT MITOCHONDRION.
FT CHAIN 1 21
FT VARIANT 22 439
FT VARIANT 6 6
FT VARIANT 40 41
FT VARIANT 49 49
FT VARIANT 58 58
FT VARIANT 82 82
FT VARIANT 119 119
FT VARIANT 137 137
FT VARIANT 140 140
FT VARIANT 147 147
FT VARIANT 189 189
FT VARIANT 192 192
FT VARIANT 206 206
FT VARIANT 212 212
FT VARIANT 231 231
FT VARIANT 267 267
FT VARIANT 366 366
FT VARIANT 378 378
FT CONFLICT 219 222
SQ SEQUENCE 439 AA; 50560 MW; 309CD950D85ACBD0 CRC64;

```

```

Query Match 4.6%; Score 145; DB 1; Length 439;
Best Local Similarity 20.8%; Pred. No. 0.0076;
Matches 75; Conservative 65; Mismatches 158; Indels 62; Gaps 12.

QY QLVLPCHTNO-----RGLSVGQLLKWIDTACLSAERH-----AGCPVVAS 93
Db :::::
88 EVLLPLGTDPRLDKYVTVQNTVVRGRILDDLSGLVLCVWHNNHNSHKYKSPLSIVVL 147
QY 94 MDDI-YEHHTISVGQVNNIKAKVNFANFSMEVGIQV--ASEDLCEKQMNVCALATPV 150
Db :::::
148 VDKIDMCGHSLSPEQDIFKTHGVSWGVTSMVEKKMKQLNHD--EKYWPVLDTAFVMV 204
QY 151 ARREITK-VXLKQTP--RTEEBKMEHSVAAERBRMELVYADTKIDLALNCAIGDLES 206
Db 205 ARDSENGKPAFVNFILPENKEBEELFKQGLINKSRIRAFSTSLK-----VAPSSSE 257
QY 207 RDC-----SRMP-----AKETVESVEVLVPRPHANQGTGGQJMA 244
Db 258 RNIITHELFTLLDPKTI SFQSRILRPPKAVWMEDTLKSLDICHQPERVVFRIIRPGFLMR 317
QY 245 WMEVAVITAA SRLCRAHEPTLKAIEKHFHGRGSQVGDRLVKLAIYNNAFKHSMEVGYCEA 304
Db 318 KAYELAMATAQSFQGSREYVTVDDIMFQKPEVGVSLFLFSQVCFOTDONYIVRHSSEV 377
QY 305 YROEAETRRRIINSAFMTFVVLADDPQQLPWIRPQGG---DGRRRYREASARKKILLD 361
Db 378 SSUDS---REMTNTVVFHFTMSKEVDFLIRPKYGESEMLYLDQGRHKSSMTPTVKKL 434

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 15:42:01 ; Search time 20 Seconds

(without alignments)
2919.413 Million cell updates/sec

Title: US-10-055-624B-2

Perfect score: 3181
Sequence: 1 MIONVGNHLRGLASVFSNR.....TKAIPVAKINRFGYLQDT 607

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:***
1: pir1:***
2: pir2:***
3: pir3:***
4: pir4:***

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3181	100.0	630	2 T00351	hypothetical prote
2	225	7.1	338	2 J05415	palmitoyl-CoA hydr
3	224.5	7.1	276	2 E98257	hypothetical prote
4	223.5	7.0	155	2 G72051	cytosolic acyl-CoA
5	223.5	7.0	155	2 C86572	acyl-CoA thioester
6	220	6.9	338	2 J05416	acyl-CoA hydrolase
7	218	6.9	343	2 J05416	palmitoyl-CoA hydr
8	216.5	6.8	157	2 F83749	acyl-CoA thioester
9	216.5	6.8	162	2 F83937	acyl-CoA thioester
10	214.5	6.7	160	2 F71502	probable acyl-coa
11	207.5	6.5	159	2 C81662	cytosolic acyl-CoA
12	204	6.4	167	2 A96036	probable acyl-coa
13	204	6.4	318	2 B90399	conserved acyl-coa
14	201	6.3	188	2 F82956	conserved acyl-coa
15	184	5.8	438	2 H84711	hypothetical prote
16	180.5	5.7	166	2 E83309	conserved hypotet
17	179	5.6	160	2 G81079	acyl CoA thioester
18	179	5.6	160	2 F81864	probable acyl-CoA
19	177	5.6	174	2 T09451	vldl protein - Hel
20	177	5.6	174	2 C64631	conserved hypotet
21	174.5	5.5	365	2 E72725	probable acyl-coa
22	174	5.5	172	2 AB1690	acyl-CoA hydrolase
23	173	5.4	176	2 AB1318	acyl-CoA hydrolase
24	170.5	5.4	172	2 E89975	conserved hypotet
25	163.5	5.1	222	2 C75539	conserved hypotet
26	157	4.9	155	2 H84243	acyl-CoA hydrolase
27	151	4.7	147	2 G75629	conserved hypotet
28	148	4.7	141	2 A83178	hypothetical prote
29	144.5	4.5	136	2 E69856	acyl-CoA hydrolase

30	142.5	4.5	158	2 AF3027	acyl-CoA hydrolase
31	133	4.2	137	2 B81365	probable hydrolase
32	133	4.2	137	2 B81365	acyl-CoA thioester
33	130.5	4.1	162	2 D82397	acyl-CoA thioester
34	124	3.9	147	2 D84393	hypothetical prote
35	123	3.9	393	2 T25908	chromodomain-helic
36	121	3.8	1711	1 A47392	MLN 64 protein - h
37	118	3.7	445	2 I38027	probable acyl-CoA
38	114.5	3.6	148	2 E81878	probable acyl-CoA
39	114	3.6	154	2 A64097	conserved hypotet
40	114	3.6	1275	2 T37192	nebulin-related pr
41	112	3.5	461	2 T30073	hypothetical prote
42	111.5	3.5	440	2 T32085	hypothetical prote
43	111	3.5	150	2 C98254	probable acyl-CoA
44	110	3.5	152	2 E87272	hypothetical prote
45	108.5	3.4	130	2 AE2821	acyl-CoA hydrolase

ALIGNMENTS

Query Match	100.0%;	Score 3181;	DB 2;	Length 630;
Best Local Similarity	100.0%;	Pred. No. 4e-238;		
Matches 607;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MIONVGNHLRGLASVFSNR	SRKSLRAGNDSAMADGEGYRNPTTEVOMQLVLPCHTNQ	60
DB	24	MIONVGNHLRGLASVFSNR	SRKSLRAGNDSAMADGEGYRNPTTEVOMQLVLPCHTNQ	83
QY	61	RGLSVGQLIKWIDTTACLSAERHAGCPVTASMDIYFEHTISVGVVNIKAKVNAFN	120	
DB	84	RGLSVGQLIKWIDTTACLSAERHAGCPVTASMDIYFEHTISVGVVNIKAKVNAFN	143	
QY	121	SSMEVGIOVASBDCSEKQNVCKALATFPAAREITRKVKQITTPREEKMHSAVER	180	
DB	144	SSMEVGIOVASBDCSEKQNVCKALATFPAAREITRKVKQITTPREEKMHSAVER	203	
QY	181	RRMRLVYADTIKDLANCAIQGDLSESDCSRMVPAEKTRVESYELVLPHPANQNTFGG	240	
DB	204	RRMRLVYADTIKDLANCAIQGDLSESDCSRMVPAEKTRVESYELVLPHPANQNTFGG	263	
QY	241	QIYAMMENVAITIASRLCRAHPTLKAIEHFHFGPSQVSGDRLYLKAIUNNAFHSMEVG	300	
DB	264	QIYAMMENVAITIASRLCRAHPTLKAIEHFHFGPSQVSGDRLYLKAIUNNAFHSMEVG	323	
QY	301	CVSAVYQEAETRRHRHNSAFMTFVLDADQPOLLPPIROPDGEERYEASARKKIRL	360	
DB	324	CVSAVYQEAETRRHRHNSAFMTFVLDADQPOLLPPIROPDGEERYEASARKKIRL	383	
QY	361	DRKYIVSCQTEVPLSVPMPSNQVLYSYNNVSLKMLYAKDNWVLSSEISQVRLYTLTD	420	
DB	384	DRKYIVSCQTEVPLSVPMPSNQVLYSYNNVSLKMLYAKDNWVLSSEISQVRLYTLTD	443	
QY	421	DKFLSFHMEVNVAVHDAQAFLILSDLRQRPBMDKHRSVELVQGVDEDDAIYHTSPALG	480	

Db 444 DKFSFHEMVAHDAQAFLILSDLRPEMDKHYRSEVELVOQVDEDAIYHTSPALG 503
 QY 481 GHTYKQDFVILASRRKPCDNDPVYIALRSVTLPTHRETPYRGETLCSGFCIMREGDQ 540
 Db 504 GHTYKQDFVILASRRKPCDNDPVYIALRSVTLPTHRETPYRGETLCSGFCIMREGDQ 563
 QY 541 LTKCQWVSLTEHVSAGFSYWGIESKRSRSGNMKLAGHSTLTKAIPVAKINSR 600
 Db 564 LTKCQWVSLTEHVSAGFSYWGIESKRSRSGNMKLAGHSTLTKAIPVAKINSR 623
 QY 601 FGVLQDT 607
 Db 624 FGVLQDT 630

RESULT 2

JC5415
 palmitoyl-CoA hydrolase (EC 3.1.2.2) brain type - rat
 N:Alternate names: long-chain fatty-acyl-CoA hydrolase
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 10-Jun-1997 #sequence_revision 12-Sep-1997 #text_change 11-Aug-2003
 C:Accession: J05415; PC4318
 R:Yamada, Y.; Furuhata, T.; Iida, N.; Watanabe, T.; Hosokawa, M.; Satoh, T.; Someya, A.; Biochem. Biophys. Res. Commun. 232, 198-203, 1997
 A:Title: Molecular cloning and expression of cDNAs encoding rat brain and liver cytosolic A:Reference number: J05415; MUID:97236308; PMID:9125130
 A:Accession: J05415
 A:Molecule type: mRNA
 A:Residues: 1-338 <YAM1>
 A:Cross-references: DDBJ:D88890; NID:G1944427; PIDN:BAA19626.1; PID:G1944428
 A:Experimental source: brain
 A:Accession: PC4318
 A:Molecule type: protein
 A:Residues: 17-29;213-221;272-300;327-337 <YAM2>
 A:Experimental source: brain
 C:Comment: This enzyme catalyzes the hydrolysis of fatty acyl-CoA thioesters to free fatty acids in intracellular protein transport, activation of protein kinase and regulation of Superfamily: acyl-CoA thioesterase
 C:Keywords: coenzyme A; cholesterol hydrolase
 F:314-330/Region: nuclear location signal

Query Match 7.1%; Score 225; DB 2; Length 338;
 Best Local Similarity 22.4%; Pred. No. 1.1e-09;
 Matches 74; Conservative 66; Mismatches 155; Indels 36; Gaps 7;
 QY 44 PTEVOMQQLVPCHTNORGBLSVQQLKWIPTTACLSAERHAG-----CPCTVASMDDI 97
 Db 9 PAALQICRIMRPDANVAGVHGVTILKMEIAGVILSTRHCNSQNGERCVAALARVERT 68
 QY 98 YFEHTISVGVVNIKAKNRAFNSMVEGIOVASEDLCEKQNVCKALAFV--ARREI 155
 Db 69 DFLSPWICGEVAHVAELTYTSKSHSVEVQVHLSNLTGKKLTNKAATLWVPLSLKNV 128
 QY 156 TKV-----KLKQITRTTEEKMEHVAERERRRLVY--ADTIKDLNCAIQDLESRDC 209
 Db 129 DKVLEVPPIVYLRQEEBEGKRYEAKQLEKEMTKRMNGDLYOQILN----- 175
 QY 210 SRWVPAEKT---RVSVELVLPPhANEQNTFGQIQIMAMENVATTASGLCRAPHTLK 265
 Db 176 ---PENTVSYSSSLIHVGPDCITLHGFGVGVTKMKDEVAAGIYAARHCKTNIVTA 231
 QY 266 AIEHFFHRRGSPQVQDRLVLAIVNNAFSGHMEVGCVEAYQOEAFTHRRH-INSAPMTFV 324
 Db 232 SVDAINPHDKIRKGCVITISGRMFTSNKSEIEVLVDADVPVNSQKRYAASAFPTVY 291
 QY 325 VLDADQPOLPLWTRPQGDGERRRYEASAR 355
 Db 292 SLNDEGKPLPVPOLVPEDEDEKRFEEGKGR 322

RESULT 3

E98257

hypothetical protein AGR_L_2016 [imported] - Agrobacterium tumefaciens (strain C58, Cerec C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
 C:Accession: E98257
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: E98257
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-276 <AGR>
 A:Cross-references: GB:AE007870; PIDN:AAK9583.1; PID:G15159470; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L_2016
 A:Map position: linear chromosome

Query Match 7.1%; Score 224.5; DB 2; Length 276;
 Best Local Similarity 25.2%; Pred. No. 9.4e-10;
 Matches 72; Conservative 42; Mismatches 127; Indels 45; Gaps 5;
 QY 48 QMSQVLPCHTNORGBLSVQQLKWIPTTACLSAERHAGCPVYASMDIYFEHTISVQ 107
 Db 18 RLIDIVFPGDINHHTLFGGTGLAMRVATIAFRFTPTVATASCRIDFRQPARIGH 77
 QY 108 VVNIAKAKNRAFNSMVEGIOVASEDLCEKQNVCKALAFVAREITKVKLKOITRT 167
 Db 78 IVEFTAPVPVKARRSLTVEVEVMAETIIGRQOHCITRGIIFMVVA-----IPEG 125
 QY 168 EEEKMEHVAERERRMRLVADTKDLLNCAIQDLESRDCSMRVPAEKTREVSVELVL 227
 Db 126 ED-----NASTVLPDLITET-----PPSPAVTWELVIF 155
 QY 228 PPHANQNTFGQIMAMENVATTASRLCRAPHTLKALIMFHRGSPQVQDRLVLAIV 287
 Db 156 PDQANSAGRMGGBAIVMTKAAEVASRYCGKLVLASBRIDPARAIBIGEIVDAQAH 215
 QY 288 VNNAFKSHSMEVGCVEAYRQAEFTHRRHN-SAPMTFVLDADDP 332
 Db 216 VERVGRSSMSIQ--TKLMSENLITGERHITATGHTVAVADKDRP 259

RESULT 4

G72051
 cytosolic acyl-CoA thioester hydrolase family protein CP0093 [imported] - Chlamydia pneumoniae
 N:Alternate names: acyl-CoA thioesterase
 C:Species: Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: G72051; A81615
 R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606; PMID:10192398
 A:Accession: G72051
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-155 <AKN>
 A:Cross-references: GB:AE001648; GB:AE001363; NID:G4376946; PIDN:AAD18793.1; PID:G4376955;
 A:Experimental source: strain CWL029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gall, S.R.; Heidelberg, J.F.; White, O.; Hickey, I.; C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255; PMID:10684935
 A:Accession: A81615
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-155 <REA>
 A:Cross-references: GB:AE002172; GB:AE002161; NID:G7189018; PIDN:AAJ37977.1; PID:G7189027;
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: yciA; CP0093


```

Query Match      7.0%; Score 223.5; DB 2; Length 155;
Best Local Similarity 35.5%; Pred. No. 4.9e-10;
Matches 50; Conservative 29; Mismatches 55; Indels 7; Gaps 2;

QY      226 VLPEPHANHGNTGGQIIMAMWENVATTAAASRLCPAHPITLKAILEMHPFGPSGVGRILVLYK 285
       : | | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      19 IFPNNDIANNTVTGGLMSLDLRLLALVAERHTESVCYTAFVDALRFYAPAYMGENTLICK 78
       | | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      286 AIVNNAFKISMEVGCVCEA---YRQEAETRRHHINSAPMTFVVLDDADDQPOLLPIWRPOP 342
       | | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      79 AAVNRTWRISLEVGKWAENIYKQE---RRHITSAVFTFAVVEDNQPIVHQIVPET 134
       | | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      343 GDGERRYREASARKTIRIDRK 363
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      135 PEKKRRYNEDRRRQRARLELK 155
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
acyl-CoA thioesterase [imported] - Chlamydomophila pneumoniae (strain J138)
C86572
CSpecies: Chlamydomophila pneumoniae, Chlamydia pneumoniae
CDate: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
CAccession: C86572
R.Shitrali, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
Article: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A.Reference number: AB6491; PMID:20330349; PMID:10871362
A.Accession: C86572
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-155 <STO>
ACross-references: GB:BAC00008; NID:g8979026; PIDN:BAA98861.1; GSPDB:GN00142
A.Experimental source: strain J138
C.Genetics:
A.Gene: ycia

Query Match      7.0%; Score 223.5; DB 2; Length 155;
Best Local Similarity 35.5%; Pred. No. 4.9e-10;
Matches 50; Conservative 29; Mismatches 55; Indels 7; Gaps 2;

QY      226 VLPEPHANHGNTGGQIIMAMWENVATTAAASRLCPAHPITLKAILEMHPFGPSGVGRILVLYK 285
       : | | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      19 IFPNNDIANNTVTGGLMSLDLRLLALVAERHTESVCYTAFVDALRFYAPAYMGENTLICK 78
       | | | | | : : : : ~ : : : : : : : : : : : : : : : : : : : : : : : :
QY      286 AIVNNAFKISMEVGCVCEA---YRQEAETRRHHINSAPMTFVVLDDADDQPOLLPIWRPOP 342
       | | | | | : : : : ~ : : : : : : : : : ~ : : : : : : : : : ~ : : : :
DB      79 AAVNRTWRISLEVGKWAENIYKQE---RRHITSAVFTFAVVEDNQPIVHQIVPET 134
       | | | | | : : : : ~ : : : : ~ : : : : ~ : : : : ~ : : : : ~ : : : :
QY      343 GDGERRYREASARKTIRIDRK 363
       : : : : : : : : : ~ : : : : ~ : : : : ~ : : : : ~ : : : : ~ : : :
DB      135 PEKKRRYNEDRRRQRARLELK 155
       : : : : : : : : : ~ : : : : ~ : : : : ~ : : : : ~ : : : : ~ : : :

RESULT 6
JCT7161
acyl-CoA hydroxylase (EC 3.1.2.20), long-chain brain specific - human
CSpecies: Homo sapiens (man)
CDate: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 11-Aug-2003
CAccession: JCT7161
R.Yamada, J.; Kurata, A.; Hirata, M.; Taniguchi, T.; Takama, H.; Furuhata, T.; Shiratori
J. Biochem. 126, 1013-1019, 1999
A.Title: Purification, molecular cloning, and genomic organization of human brain long-c
A.Reference number: JCT7161; PMID:20047053; PMID:10578051
CAccession: JCT7161
A.Molecule type: mRNA
A.Residues: 1-338 <YAM>
ACross-references: DDB:D88894; NID:g6580075; PIDN:BAA24350.1; PID:g2780414
A.Experimental source: brain cytosol
CComment: This enzyme catalyzes the hydrolysis of fatty acyl-coAs to the corresponding
the cellular concentrations of acyl-CoA and serves as a scavenger of free long-chain acy
C.Genetics:
A.Gene: bach

```

[illegible]

Db 181 -PEPTVSYSSQSLIHLVGPDSCTLHGFGVGTMKLMDVEAGVIAARHCKINTIVTASVD 239
CY 269 MFHFRGSPQVQDRLVLKAIYNNAFKHSMEVGVCEAYRQEAETHRRH-INSAPMTFVLDDADDP 327
Db 240 AVNFHDKIRKCCVLTISGRMPTSNKSMIEVLVDADPVNDSQRYRAASAFFTYVSLN 299
CY 328 ADDQPOLLPWTRPOPGRYREASARKIRLDR 355
Db 300 QEKRPVLPVQVLPTEDEKKRFEKGR 327

RESULT 8

F83749
acyl-CoA hydrolase BH0798 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: F83749
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: F83749
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-157 <STD>
A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04517.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0798

Query Match 6.8%; Score 216.5; DB 2; Length 157;
Best Local Similarity 32.0%; Pred. No. 1.7e-09;
Matches 48; Conservative 32; Mismatches 67; Indels 3; Gaps 2;

CY 214 PAEKTRESEVLVLPPEAHNQGNTFGQIMAMMENAVTIAASRLCRAHPTLKAIEMFHR 273
Db 6 PVERSRITQTRVLVPPDTNHLGITFGGKVLAYIDEIAALTMKANSAYVTASIDSVDFK 65
CY 274 GPSQVGDRLVLKAIYNNAFKHSMEVGVCEAYRQEAETHRRHNS-APMTFVLDDADDP 332
Db 66 SSATVGDALHEGVTTHGRISME--VYRVHSSNNLTGERTLTTESPFLTVAVADESGK 123
CY 333 QLLPWIRPOPGRYREASARKIRLDR 362
Db 124 KPVPQVEPQTEEKRLVETAPARKENRKR 153

RESULT 9

F83937
acyl-CoA thioester hydrolase BH2302 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: F83937
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: F83937
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-162 <STD>
A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06021.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2302

Query Match 6.8%; Score 216; DB 2; Length 162;
Best Local Similarity 35.1%; Pred. No. 2e-09;
Matches 53; Conservative 31; Mismatches 63; Indels 4; Gaps 3;

CY 214 PAEKTRESEVLVLPPEAHNQGNTFGQIMAMMENAVTIAASRLCRAHPTLKAIEMFHR 273
Db 7 PSESLTWT-DIVLPDTNHYGTIFGGVAVYDKVASIAAMRCKRQVVTASSDSLDFI 65

CY 274 GPSQVGDRLVLKAIYNNAFKHSMEVGVCEAYRQEAETHRRHIN-SAPMTFVLDDADDP 332
Db 66 SPIRTGEALCLEGIVTTRKTSMEVFRVEA--EDLTLGERRLTATSYTLVVALNPDGK 123
CY 333 QLLPWIRPOPGRYREASARKIRLDR 363
Db 124 AEIPPVVPTTEEEKMHWGAKERYEIRRR 154

RESULT 10

F71502
probable acyl-coa thioesterase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: F71502
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tract
A:Reference number: A71570; MUID:99000809; PMID:9784136
A:Accession: F71502
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <ARN>
A:Cross-references: GB:AE001324; GB:AE001273; NID:g3328966; PIDN:AAC68137.1; PID:g332897;
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: ycjA

Query Match 6.7%; Score 214.5; DB 2; Length 160;
Best Local Similarity 34.1%; Pred. No. 2.6e-09;
Matches 47; Conservative 30; Mismatches 54; Indels 7; Gaps 2;

CY 226 VLPPHANHGNTFGQIMAMMENAVTIAASRLCRAHPTLKAIEMFHRGSPQVQDRLVLK 285
Db 25 IFPNDLNTNNTIFGGLMSLDRLLVVAERHCESICVTAFVDAMRFYAPAYMGENTLICC 84
CY 286 AIYNNAFKHSMEVGVCEA--YRQEAETHRRHNSAPMTFVLDDADDPQLPWIRPOP 342
Db 85 ASVRSKRTSLVGVKVAENITKOE---HRHITSALTFTVAADKNSPVEBELIDES 140
CY 343 GDGRRYREASARKIRL 360
Db 141 QEIRRRFRPADQRRALRL 158

RESULT 11

C81662
cytosolic acyl-CoA thioester hydrolase family protein TC0822 [imported] - Chlamydia muric
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: C81662
R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, F.
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10664935
A:Accession: C81662
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <ITEN>
A:Cross-references: GB:AE002348; GB:AE002160; NID:g7190839; PIDN:AAF39624.1; PID:g719085C
A:Experimental source: strain Ni9g (MoPn)
C:Genetics:
A:Gene: TC0822

Query Match 6.5%; Score 207.5; DB 2; Length 159;
Best Local Similarity 33.3%; Pred. No. 8.8e-09;
Matches 46; Conservative 30; Mismatches 55; Indels 7; Gaps 2;

CY 226 VLPPHANHGNTFGQIMAMMENAVTIAASRLCRAHPTLKAIEMFHRGSPQVQDRLVLK 285
Db 24 IFPNDLNTNNTIFGGLMSLDRLLVVAERHCESICVTALVDVAFYAPAYMGENTLICC 83

```
QY      286 AIVNNAFKSHMEVGVCEA---YRQEAETRRHINSAFMTFVLADDDQPOLMPWRPQ 342
      84 ASVNRSWRTSLSEVGVKWMANITYKQ-----QRHTSATYTFVAVDENNSPEIPELVET 139
Db
QY      343 GDGERRRYREASARKKRL 360
      140 QEEIRRFHEADQRASRL 157
Db

RESULT 12
A96036
Probable acyl-CoA thioester hydrolase protein (EC 3.1.2.-) [imported] - Sinorhizobium me
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: A96036
R:Finan, T.M.; Weidner, S.; Wong, K.; Bhurmesler, J.; Chain, P.; Vorholter, F.J.; Hernar
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: A96036
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1167 <KOR>
A:Cross-references: GB:AL591985; PID:CAC49953.1; PID:G15141441; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Gallibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
Pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaurie,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMD21634
A:Genome: plasmid
C:Keywords: thioester hydrolase

Query Match      6.4%; Score 204; DB 2; Length 167;
Best Local Similarity 32.2%; Pred. No. 1.8e-08;
Matches 48; Conservative 31; Mismatches 68; Indels 2; Gaps 2;

QY      212 MVEPAEKTR-VESVELVLPPIANHQGTFGGQIMAMENVAITIASRLCRAHPTLKAIEMF 270
      1 MMTGEQTRHLVMTVLTMPDMANPSGKXHGALINLDRVAFSCASRSKQYAVTISVDQV 60
Db
QY      271 HFRGPGSQVDRLVKAIVNNAFKSHMEVGVCEAYRQEAETRRHINSAFMTFVLADDD 330
      61 IFKEPIHVGELVTFRAINYAGRTSMVGIRVEAKENIRAGT-RRHTNSCYFTWVAVDAG 119
Db
QY      331 QPOLPWRPQPDGGERRRYREASARKKIR 359
      120 RPTVPEYHPETATKERRHQAAELRLRL 148
Db

RESULT 13
B90399
Probable acyl-coenzyme A thioester hydrolase [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 11-Aug-2003
C:Accession: B90399
R:She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.
arct, K.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
Submitted to Genbank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: B90399
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-318 <KOR>
```

```
A:Cross-references: GB:AE006641; NID:G13815589; PIDN:AAK42449.1; GSPDB:GN00155
C:Genetics:
A:Gene: S902287
C:Superfamily: acyl-CoA thioesterase

Query Match      6.4%; Score 204; DB 2; Length 318;
Best Local Similarity 22.0%; Pred. No. 4.4e-08;
Matches 72; Conservative 67; Mismatches 132; Indels 56; Gaps 9;

QY      58 TNGRGELSVGLKMDITTTACLSAEHRHAGCPCTASMDIYFPHITISVGVVNIKAVNR 117
      22 SNFLNRLHGGDMPLFVETMTISASKVAMTTILASIDVDFKKPVRIGDIVKRAATVY 81
Db
QY      118 AFNSHMEVGIQVASEDLCEKQNVCKALATFYARREI---TVKYLQITPREEEKMEH 174
      82 IGMTSLSEVEYRARPVD-----EVSVAVTYVQVDLRLPTPNVR-IMASEDDRXI 134
Db
QY      175 SVAAERRRML-----VYADTKDLANCAIOGDESRCSMVFAEKTRVSYE 224
      135 DEAKKRREMLSKIVNRQKKRFYVDDITDGL-----RRYISNV- 172
Db
QY      225 LVLPPIANHQGTF-GGQIMAMENVAITIASRLCRAHPT-----LKAIEMHF 272
      173 IHVSPELTYDGRIMSACKLKMDDLGLICLRLDHNSNLIDNSFNNAVTVAVKGLAF 232
Db
QY      273 RGPQVQDRLVLAIVNNAFKSHMEVGVCEAYRQEAETRRHINSAFMTFVLADDDQ 332
      233 YSPLRLNDILIRA--GLVVGNTSADILINVRREDLNGAKEHVAATVFTYVRVDKQKP 290
Db
QY      333 QOLPWRPQPDGGERRRYREASARKKIR 359
      291 IKIPEYPTVERERKRLVEELTRGLR 317
Db

RESULT 14
F82956
conserved hypothetical protein PA5519 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F82956
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mitsuuchi, S.D.; Warriner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lim.
.; Loiy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: F82956
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <STO>
A:Cross-references: GB:AB004964; GB:AB004091; NID:G9951849; PIDN:AA089904.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA5519

Query Match      6.3%; Score 201; DB 2; Length 188;
Best Local Similarity 30.2%; Pred. No. 3.3e-08;
Matches 54; Conservative 36; Mismatches 69; Indels 20; Gaps 3;

QY      13 LASVFSNRKS-----RKSALRAGNDSAMADGEGYRPTVEVMSQVLYLPCHITNRGHSV 66
      9 LPSEFASVNSPARPDERKSAQPGNN-----QLSMVLTMPDMANPSGAVHG 55
Db
QY      67 GOLLKMDITTAASAERHAGCPCTASMDIYFPHITISVGVVNIKAVNRAFNSSMEVG 126
      56 GTLIXYIDEVAIYACASRYAGRYVTTLSVDQVIFRPHVGEIVTFLASVNTGTSMEIG 115
Db
QY      127 IOVASEDLCEKQNVCKALATFYARREITK-VKLOITPREEEKMEHSAARRRR 184
      116 VKVITTEIRBQSVRHTNSCFETWVALDDEKRPVAVRLEJETABQKRFRQAQQRQLR 174
Db

RESULT 15
```

H84711
hypothetical protein At2g30720 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 11-Aug-2003
C:Accession: H84711
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eilen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: H84711
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-438 <STO>
A:Cross-references: GB:AB02093; NID:g2880049; PIDN:AAC02743.1; GSPDB:GND0139
A:Genetics:
A:Gene: At2g30720
A:Map position: 2
C:Superfamily: acyl-CoA thioesterase

Query Match 5.8%; Score 184; DB 2; Length 438;

Best local Similarity 25.0%; Pred. No. 2.5e-06;
Matches 96; Conservative 51; Mismatches 171; Indels 66; Gaps 13;

```
QY 39 EGYRN---PTEVQMSQLVPC-----HTNQGELSVGOLLKMWIDTTACLSEAERHAG-- 86
Db 77 EARNRMFEIPTGDDASQSKLTAKSPSRRTNPNWNEIRTKGLVEDDALAGTISFKHCGD 136
QY 87 -----CPCVTASMDIYEHTISVGQVNIKAKVNRANFNSSMEVGIOVAS-EDLCSEKQW 140
Db 137 SSARSMILVTASVDRIIMKRPFRVDYDLSIVGAVTWGRSSMEMQVLIQIOTNNSSES 196
QY 141 NVCKALMEFVARREIT--KVLKQITPREEEKMEHSVAERRRMR-----L 185
Db 197 VALBNFTFVARDOQTGSAPIINOVPETEHKEKIMKEAERNNKIKOKRAQGEHEKEL 256
QY 186 VYADTIKDLANCAIQDLESRDGSRMVPAREKTVESVELVLPHPANHQNTFGGOIMAW 245
Db 257 KDLERINELLAGRVEFLMPLADNLSILIKDISHENSLICQPOQRNTHGRIFGFTMRK 316
QY 246 MENVATIASRLCRAHPTLKAIEMFHRGSPQVGDRLVKA-IVNNAFKHSME--VGVCV 302
Db 317 AFELAFSNAVYTFAGVSPRFLEVDVDFIKPVDVGNFLRFKSRVLYTEATSSAEPLINIEV 376
QY 303 EAYRQEAETHRRHNSAF-MTFVVLADADQQLBFWIRPQGDGERRRY-----EASAR 355
Db 377 VAHVTSPELRSESVSNRPYFTFSV-----RPEAMKDGKLRNVVPATEEBAR 423
QY 356 KKI-RLDRKYIVSCQTEVPLSV 378
Db 424 RVIERM-----AERPISLP 438
```

Search completed: August 12, 2004, 15:45:52
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 15:40:21 ; Search time 58 Seconds
(without alignments)
2957.006 Million cell updates/sec

Title: US-10-055-624B-2

Perfect score: 3181
Sequence: 1 MTQNVGNHLRRGLAVFSNR.....TLKALPARKINSRFGYLODT 607

Scoring table: BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Geneseqp29t9n04:*
2: geneseqp1980s:*
3: geneseqp1990s:*
4: geneseqp2000s:*
5: geneseqp2001s:*
6: geneseqp2002s:*
7: geneseqp2003as:*
8: geneseqp2003bs:*
9: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2838	89.2	594	4	AAU77063 Human pro
2	970	30.5	246	4	AAU18380 Human end
3	956	30.1	207	4	AAU20561 Human sec
4	956	30.1	207	4	AAU18487 Human end
5	413	13.0	119	3	AAU58891 Breast an
6	255	8.0	339	4	AAU76624 Corynebac
7	255	8.0	339	4	AAU92590 C glutam
8	225	7.1	338	7	AAU56432 Rat Prote
9	224.5	7.0	371	5	AAU69844 Human pol
10	223.5	6.9	155	2	AAU5275 Chlamydia
11	223.5	6.9	155	2	AAU6796 Protein e
12	220	6.9	339	6	AAU74896 Human sec
13	220	6.9	339	5	AAU93448 Human nov
14	220	6.9	339	6	AAU5399 Region of
15	219	6.9	370	6	AAU5399 Human pro
16	218.5	6.9	163	2	AAU37682 Amino aci
17	207.5	6.5	159	6	AAU27389 Protein e
18	206.5	6.5	375	3	AAU71115 Human Hyd
19	203	6.4	168	6	AAU17829 Protein e
20	191.5	6.0	172	6	AAU4404 Actinobac
21	188	5.9	389	4	AAU3806 Novel hum
22	186.5	5.9	177	5	AAU3851 Staphyloc
23	186	5.8	156	6	AAU1697 Protein e
24	186	5.8	370	6	AAU1179 Human DIT
25	185	5.8	176	6	AAU42699 Protein e

26	182	5.7	455	5	AAU77063
27	181.5	5.7	183	6	AAU35445
28	181	5.7	160	6	AAU79796
29	179	5.6	389	4	AAU58313
30	177	5.6	174	6	AAU30905
31	175.5	5.5	404	4	AAU6274
32	173	5.4	172	5	AAU48981
33	173	5.4	172	6	AAU32543
34	171.5	5.4	170	6	AAU17573
35	171.5	5.4	177	6	AAU3445
36	170.5	5.4	176	6	AAU2462
37	170.5	5.4	181	6	AAU71499
38	170.5	5.4	472	4	AAU58527
39	166	5.2	178	6	AAU28987
40	148	4.7	385	4	AAU58316
41	145	4.6	439	2	AAU74418
42	142	4.5	448	3	AAU71118
43	140.5	4.4	438	5	AAU77062
44	140	4.4	137	6	AAU22724
45	139.5	4.4	131	6	AAU21871

ALIGNMENTS

RESULT 1
AAU95601 standard; protein; 594 AA.

ID AAU95601: (first entry)

DE Human protein sequence SEQ ID NO:18290.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-MAY-2000; 2000JP-00183767.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto T;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

XX diagnosis of the abnormality of the proteins encoded by the full-length

XX cDNAs.

XX Claim 8; SEQ ID NO 18290; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-

XX length cDNAs defined in the specification, where a primer set comprises:

XX (a) an oligo-dT primer and an oligonucleotide complementary to the

XX complementary strand of a polynucleotide which comprises one of the 5602

XX nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

SO Sequence 594 AA:

Query Match 89.2%; Score 2838; DB 4; Length 594;
Best Local Similarity 100.0%; Pred. No. 6,9e-278;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIVGVGHLRGLASVPSNRKTSKALPAGNDSAMAOEGFRNPTVOMSQLVLPCHNQ 60
DB 1 MIVGVGHLRGLASVPSNRKTSKALPAGNDSAMAOEGFRNPTVOMSQLVLPCHNQ 60
QY 61 RGEISVGLLKMIDTTACTAERHAGCPCTASMDIYFEHTISVGQVNIKAKVNAFN 120
DB 61 RGEISVGLLKMIDTTACTAERHAGCPCTASMDIYFEHTISVGQVNIKAKVNAFN 120
QY 121 SSMVEGIVASEDCISEKQNVCKALATFVARREITVKYKQITPRTEEMHESVAER 180
DB 121 SSMVEGIVASEDCISEKQNVCKALATFVARREITVKYKQITPRTEEMHESVAER 180
QY 121 SSMVEGIVASEDCISEKQNVCKALATFVARREITVKYKQITPRTEEMHESVAER 180
DB 121 SSMVEGIVASEDCISEKQNVCKALATFVARREITVKYKQITPRTEEMHESVAER 180
QY 181 RRMPLVYADTKDILLANCAIQDLESRDSCRMVPEKTRVSEVLVLPphanHOGNTFGG 240
DB 181 RRMPLVYADTKDILLANCAIQDLESRDSCRMVPEKTRVSEVLVLPphanHOGNTFGG 240
QY 241 QIMAMMENVAITIASRLCRAPTKALIMPFRRGSPQVGDILVKALVNNAFKSMMEGV 300
DB 241 QIMAMMENVAITIASRLCRAPTKALIMPFRRGSPQVGDILVKALVNNAFKSMMEGV 300
QY 241 QIMAMMENVAITIASRLCRAPTKALIMPFRRGSPQVGDILVKALVNNAFKSMMEGV 300
DB 241 QIMAMMENVAITIASRLCRAPTKALIMPFRRGSPQVGDILVKALVNNAFKSMMEGV 300
QY 301 CVEAYRQEAETRRHINSAPMTFVLADDOQPLLPIWIRPQGGERRRYSARKKIRL 360
DB 301 CVEAYRQEAETRRHINSAPMTFVLADDOQPLLPIWIRPQGGERRRYSARKKIRL 360
QY 361 DRKIVVSCQTEVPLSVWDPDSNOVYLSYNNVSSIKMLVAQDNVLSSEISQVRLYTLED 420
DB 361 DRKIVVSCQTEVPLSVWDPDSNOVYLSYNNVSSIKMLVAQDNVLSSEISQVRLYTLED 420
QY 421 DKPLSFHMEVYVHDAQAQAFLLSDLRPEMDKHYRSEVSVQVDEBDATYHTSPALG 480
DB 421 DKPLSFHMEVYVHDAQAQAFLLSDLRPEMDKHYRSEVSVQVDEBDATYHTSPALG 480
QY 481 GHTYRQDFVLIASRRKPCDNGDPYIALRSVTLPTHRETPYRGETLCSGFCIMREGDQ 540
DB 481 GHTYRQDFVLIASRRKPCDNGDPYIALRSVTLPTHRETPYRGETLCSGFCIMREGDQ 540
QY 541 LTK 543
DB 541 LTK 543
QY 541 LTK 543
DB 541 LTK 543

RESULT 2
AAU18380 standard; protein, 246 AA.
AAU18380;
AAU18380;
21-NOV-2001 (first entry)
Human endocrine polypeptide SEQ ID No 335.
Endocrine protein; human; mouse; rabbit; goat; horse; food additive; cat;

KW dog; chicken; sheep; immunosuppressive; antiarthritis; vasotropic;
KW antineoplastic; antiproliferative; cytostatic; cardiac; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; viricide; fungicide; cancer;
KW ophthalmological; vinnery; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-infertility.

OS Homo sapiens.
XX
XX MO20015364-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001308.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215155P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226819P.
XX 22-AUG-2000; 2000US-0226868P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.

DT 06-DEC-2001 (first entry)
DE Human secreted protein, Seq ID No 553.
XX
XX Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW Rheumatoid arthritis; antiarteriosclerotic; cardiac; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing.
XX
XX Homo sapiens.
XX WO200155326-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001347.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-MAR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205151P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216547P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226688P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
PR

PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236387P.
PR 29-SEP-2000; 2000US-0236388P.
PR 29-SEP-2000; 2000US-0236399P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR

PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI, 2001-451931/48.
XX N-PSDB; AAS33270.
DR
XX
XX New nucleic acids and polypeptides, useful for diagnosing, preventing or
PT treating medical conditions.
PT
PS Claim 11, SEQ ID NO 553, 753bp, English.

```
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241222P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250161P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251969P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-451936/48.
XX N-PSDB; AAS29716.
XX
XX Isolated polypeptide for treating, preventing and/or diagnosing
PT disorders of the endocrine system such as reproductive disorders,
PT endocrine cancers and also for testing and detection e.g. diagnosis.
XX
XX Claim 11; SEQ ID NO 442; 604pp; English.
XX
XX Sequences AAU18282-AAU18507 represent endocrine polypeptides of the
CC invention. Endocrine polypeptides and their associated polynucleotides
CC are useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC determining the presence or absence of a mutation in an endocrine
CC polynucleotide. The treatable disorders include autoimmune diseases such
CC as rheumatoid arthritis, hyperproliferative disorders such as neoplasms
CC of the breast or liver, cardiovascular disorders such as cardiac arrest,
CC cerebrovascular disorders such as cerebral ischaemia, nervous system
CC disorders such as Alzheimer's disease, infections caused by bacteria,
CC viruses and fungi, ocular disorders such as corneal infection, endocrine
CC disorders such as premature labour and infertility, gastrointestinal
CC disorders such as Crohn's disease, renal disorders such as
CC glomerulonephritis and respiratory disorders such as asthma. The
CC polypeptides can also be used to aid wound healing, to prevent skin aging
CC due to sunburn, to maintain organs before transplantation, to regenerate
CC tissues and in chemotaxis. The polypeptides can also be used as a food
CC additive or preservative to increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

Query March 30.1%; Score 956; DB 4; Length 207;
Best Local Similarity 95.3%; Pred. No. 6.9e-89;
Matches 181; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 360 LDRKYIVSCQTEVPLSYWPDPSNQVLYSYNNVSLKMLVAKDMWVLSSEISQVRLTYLLE 419
DB 12 LPRKYIVSCQTEVPLSYWPDPSNQVLYSYNNVSLKMLVAKDMWVLSSEISQVRLTYLLE 71
QY 420 DDKFLSFMEVNVVHDAQAQFLILSLDFORPEMDKHYRSEVLVQVDEDDAIYHTSPAL 479
DB 72 DDKFLSFMEVNVVHDAQAQFLILSLDKOREPEMDKHYRSEVLVQVDEDDAIYHTSPAL 131
QY 480 GGHTKPODFVILASRRKPCDNGDPVILALRSVTLPTHETPEYRGETLCSGFCILMRGD 539
DB 132 GGHTKPODFVILASRRKPCDNGDPVILALRSVTLPTHETPEYRGETLCSGFCILMRGD 191
QY 540 QLTCKCWVRV 549
```

Db 192 QLTVCASRL 201

RESULT 5
AAB58891
ID AAB58891 standard; protein; 119 AA.

DT 27-MAR-2001 (first entry)

DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 599.

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
XX neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX anti-diabetic; anti-inflammatory; anti-ulcer; vulnerary; anticonvulsant;
XX antibacterial; antifungal; antiparasitic; cardiatic; immune disorder;
XX Addison's disease; allergy; autoimmune haemolytic anaemia;
XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX cardiovascular disorder; wound healing; neurological disease.

OS Homo sapiens.

PN WO200055173-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US005881.

PR 12-MAR-1999; 99US-0124270P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WEI; 2000-611515/58.

DR N-PSDB; AAF21794.

PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention, treatment
PT and diagnosis of cancer, immune disorders, cardiovascular disorders and
PT neurological diseases.

PS Claim 11; Page 1036; 12999p; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX associated with breast and ovarian cancer. Included in the invention are
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX isolation and characterisation of the DNA and protein sequences of the
XX invention. The breast and ovarian cancer associated DNA, protein, agonist
XX or antagonist sequences exhibit cytostatic; immunosuppressive; neurotropic;
XX neuroprotective; antiviral; antiallergic; hepatotropic; anti-diabetic;
XX anti-inflammatory; anti-ulcer; vulnerary; anticonvulsant; antibacterial;
XX antifungal; antiparasitic and cardiatic activity. The polynucleotide and
XX protein sequences are used in the diagnosis of cancer, particularly
XX breast and ovarian cancer. The nucleic acid sequences, proteins, agonists
XX and antagonists may also be used in the diagnosis, prevention and treatment
XX of immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
XX cardiovascular disorders such as myocardial ischemias; wound healing;
XX neurological diseases such as cerebral anoxia and epilepsy; and
XX infectious diseases

XX Sequence 119 AA;

Query Match 13.0%; Score 413; DB 3; length 119;
Best Local Similarity 96.4%; Pred. No. 3e-33;
Matches 81; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 383 NOVYLSYNNVSSIKMLVAKDNWVLSSEISQVRLTYLLEDDKFLSFHMEMVVDAAQAELL 442
Db 4 DQVYLSYNNVSSIKMLVAKDNWVLSSEISQVRLTYLLEDDKFLSFHMEMVVDAAQAELL 63

QY 443 LSDLRQRPEDMKHRSVELVQOVD 466
Db 64 LSDLRQRPEDMKHRSVELVQOVD 87

RESULT 6
AAB76624
ID AAB76624 standard; protein; 339 AA.

AC AAB76624;

DT 11-APR-2001 (first entry)

DE Corynebacterium glutamicum MCT protein SEQ ID NO:230.

XX Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;
XX membrane construction and membrane transport protein; petroleum spill;
XX hydrocarbon degradation; gram positive aerobic bacterium; marker;
XX identification; microorganism; fine chemical production; transformation;
XX genome mapping; genetic engineering.

OS Corynebacterium glutamicum.

PN WO200100805-A2.

PD 04-JUN-2001.

PF 23-JUN-2000; 2000WO-IB000926.

PR 25-JUN-1999; 99US-0141031P.

PR 08-JUL-1999; 99DE-01031454.

PR 08-JUL-1999; 99DE-01031478.

PR 08-JUL-1999; 99DE-01031563.

PR 09-JUL-1999; 99DE-01032122.

PR 09-JUL-1999; 99DE-01032124.

PR 09-JUL-1999; 99DE-01032125.

PR 09-JUL-1999; 99DE-01032128.

PR 09-JUL-1999; 99DE-01032180.

PR 09-JUL-1999; 99DE-01032182.

PR 09-JUL-1999; 99DE-01032191.

PR 09-JUL-1999; 99DE-01032209.

PR 09-JUL-1999; 99DE-01032212.

PR 09-JUL-1999; 99DE-01032227.

PR 09-JUL-1999; 99DE-01032228.

PR 09-JUL-1999; 99DE-01032229.

PR 14-JUL-1999; 99DE-01032230.

PR 14-JUL-1999; 99DE-01032927.

PR 14-JUL-1999; 99DE-01033005.

PR 14-JUL-1999; 99DE-01033006.

PR 27-AUG-1999; 99DE-01040764.

PR 27-AUG-1999; 99DE-01040765.

PR 27-AUG-1999; 99DE-01040766.

PR 27-AUG-1999; 99DE-01040810.

PR 27-AUG-1999; 99DE-01040831.

PR 27-AUG-1999; 99DE-01040832.

PR 27-AUG-1999; 99DE-01040833.

PR 31-AUG-1999; 99DE-01041378.

PR 31-AUG-1999; 99DE-01041379.

PR 31-AUG-1999; 99DE-01041395.

PR 03-SEP-1999; 99DE-01042077.

PR 03-SEP-1999; 99DE-01042078.

PR 03-SEP-1999; 99DE-01042079.

PR 03-SEP-1999; 99DE-01042088.

XX (BADI) BASF AG.

PA Pompejus M, Kroegeer B, Schroeder H, Zelder O, Heberhauer G;

XX PI

DR WPI; 2001-071486/08.
DR N-PSDB; AAF67857.
XX
XX Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
XX transformation.
PS Claim 20; Page 495-496; 1119pp; English.
XX
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAF676510 to
CC AAF76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention
XX
SQ Sequence 339 AA;

Query Match 8.0%; Score 255; DB 4; Length 339;
Best Local Similarity 25.7%; Pred. No. 1.8e-16;
Matches 89; Conservative 56; Mismatches 171; Indels 30; Gaps 10;

QY 41 YRNPLEVQMSQVLPCHTNQSGELSV--GQLKMTDTTACLSARHAGCPVTASMDIY 98
DB 14 HRSP-EVTLRPMAPRTDVLNAGSHGVGGRVLEMDKAAVACATQWSGTCVTAYVGH 72
QY 99 FEHTISVQGVNIAKAVNAFNSMEVGIQVASEDLCEKQWVCKALATVAREIT-- 156
DB 73 FTIRIPSGHMEVBSRIAMTGRSSMHIVNEVLSADPRDGNTRACDCLVITVAKDTATGR 132
QY 157 KVKLKQITPRTEERKMEHSVAERRMFLVYADITIKDLNCAIQGDESDDCSRMVPAE 216
DB 133 ATPVPSFTPKNEEQRVLEANSRIGLRK-----ALAEWEKQTYNG--PSE 177
QY 217 KTRVESVELVPPRPHANQGNFTGGQIMAMMENVAITIASRLCRAPHTLKAEMFHRGPS 276
DB 178 APRLITRFLAKPTDINNCGKXVHGATAMEMIDEAGACTMWSGNHTVAVAGIRFYQPI 237
QY 277 QVGRDLVYKAIVNNAFHSMEVGVCV--EAYROBAETHRR-HINSAPMTFVLADADQP 332
DB 238 QIGDLIEVDAMRTDKRSQMSTHVRAGDAHRGAELETAIH--ATVTYIGIDVDGEP 294
QY 333 QLLPWIRPQPGDGERRRYREASARKKIRLDRKYI--VSCQTEVPISV 377
DB 295 LPAPQFVPRTPEDIQLAEHANILRDLRAD--YTPMPLFQRRVPIQI 338

RESULT 7
AA92590
ID AA92590 standard; protein, 339 AA.
XX
XX AA92590;
AC
XX 26-SEP-2001 (first entry)
DT
XX C glutamicum protein fragment SEQ ID NO: 6344.
DE
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS
XX EPI108790-A2.
PN
XX 20-JUN-2001.
PD
XX 18-DEC-2000; 2000EP-00127688.
PF
XX

PR 16-DEC-1999; 99JP-00377484.
PR 07-APR-2000; 2000JP-00159162.
PR 03-AUG-2000; 2000JP-00280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
PI WPI; 2001-376931/40.
DR N-PSDB; AAF67809.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX
PS Claim 17; SEQ ID NO 6344; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and analyzing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described in the
CC exemplification of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the European Patent Office
XX
SQ Sequence 339 AA;

Query Match 8.0%; Score 255; DB 4; Length 339;
Best Local Similarity 25.7%; Pred. No. 1.8e-16;
Matches 89; Conservative 56; Mismatches 171; Indels 30; Gaps 10;

QY 41 YRNPLEVQMSQVLPCHTNQSGELSV--GQLKMTDTTACLSARHAGCPVTASMDIY 98
DB 14 HRSP-EVTLRPMAPRTDVLNAGSHGVGGRVLEMDKAAVACATQWSGTCVTAYVGH 72
QY 99 FEHTISVQGVNIAKAVNAFNSMEVGIQVASEDLCEKQWVCKALATVAREIT-- 156
DB 73 FTIRIPSGHMEVBSRIAMTGRSSMHIVNEVLSADPRDGNTRACDCLVITVAKDTATGR 132
QY 157 KVKLKQITPRTEERKMEHSVAERRMFLVYADITIKDLNCAIQGDESDDCSRMVPAE 216
DB 133 ATPVPSFTPKNEEQRVLEANSRIGLRK-----ALAEWEKQTYNG--PSE 177
QY 217 KTRVESVELVPPRPHANQGNFTGGQIMAMMENVAITIASRLCRAPHTLKAEMFHRGPS 276
DB 178 APRLITRFLAKPTDINNCGKXVHGATAMEMIDEAGACTMWSGNHTVAVAGIRFYQPI 237
QY 277 QVGRDLVYKAIVNNAFHSMEVGVCV--EAYROBAETHRR-HINSAPMTFVLADADQP 332
DB 238 QIGDLIEVDAMRTDKRSQMSTHVRAGDAHRGAELETAIH--ATVTYIGIDVDGEP 294
QY 333 QLLPWIRPQPGDGERRRYREASARKKIRLDRKYI--VSCQTEVPISV 377
DB 295 LPAPQFVPRTPEDIQLAEHANILRDLRAD--YTPMPLFQRRVPIQI 338

RESULT 8
ADE56432
ID ADE56432 standard; protein, 338 AA.
XX
XX ADE56432;
AC
XX 29-JAN-2004 (first entry)
DT
XX Rat Protein BAA19626, SEQ ID NO 2285.
DE
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW

CC directly from WIP0 at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 371 AA;

Query Match 7.1%; Score 224.5; DB 5; Length 371;
 Best Local Similarity 22.4%; Pred. No. 2.5e-13;
 Matches 83; Conservative 66; Mismatches 174; Indels 47; Gaps 8;

QY 9 LRRGLAVSFNRTSRKSLRAGNDSAMA---DDEGYNPTEVQMSQVLVPCHTNNGEL 64
 DB 10 LRRGLPPTC-----ALLPPASAAAAPMSGPDVETPSAIQICRMRDDANVAGNV 62
 QY 65 SVGQLLKWIDTACISAEHAG-----CPCVTASMDIYEHTISVGOVNIKAVNRA 118
 DB 63 HGGTILKMEIAGAIISTRHCSNGERCVALARVENTDLSMCGIHFVAHVAHEITYT 122
 QY 119 FNSMEVGIQVASEDLCEKQNVCKALATFVARREITVKLKQITP-----RTBEEM 172
 DB 123 SKHSVEVQVNVMSHNLTKAKKLTNKATLWVPLSLKAVDKYLEVPVYSGROQEEGR 182
 QY 173 EHSVAERRRRLVY--ADTIKDLNCAIQDLESRCSPVPAEKT---RVESVELY 226
 DB 183 KRYPAQKLERMETKRGDVIQPVLN-----PEPNTVSYSQSSLIHLV 225
 QY 227 LPPHANHQNTFGQIMAMMENAVTTIASRLCRAPTLKAIEMHFPGSPQVGRVLVKA 286
 DB 226 GPSCTLHGFPHGVTMCLMEVAGIYAAKCKTNIYASVDALNFHKIKGCYITISG 285
 QY 287 IVNNAFKHSMEVGVCEAYRQEAETHRRH-INSAPMTEFVLADDOQLLFWIRPQDG 345
 DB 286 RMTFTSNKSMIEVLVDADPVVDSQKRYRAASAFYTVLSQGRSLPVPQLVETED 345
 QY 346 ERRYREASAR 355
 DB 346 KKRFEKGKR 355

RESULT 10

AA35275
 ID AAY35275 standard; protein, 155 AA.

XX AAY35275;
 AC
 XX 17-OCT-2003 (revised)
 DT 13-SEP-1999 (first entry)
 XX
 DE Chlamydia pneumoniae transmembrane protein sequence.
 XX
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis; vaccine;
 KW neutralising epitope.

XX Chlamydia pneumoniae.
 XX OS
 XX WO9927105-A2.
 XX
 PD 03-JUN-1999.
 PF 20-NOV-1998; 98WO-IB001890.
 XX
 PR 21-NOV-1997; 97FR-00014673.
 PR 04-NOV-1998; 98US-0107078P.
 XX
 PA (BEST) GENSET.
 XX
 PI Griffais R;
 XX
 DR WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae.
 PS Page 1101; Disclosure; 1912pp; English.
 XX

CC AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY3190) of Chlamydia pneumoniae. C.
 CC pneumoniae causes respiratory disease such as pneumonia and bronchitis,
 CC and is thought to be a contributing factor in heart disease, sarcoidosis,
 CC sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The
 CC polypeptides encoded by the open reading frames of the C. pneumoniae
 CC genome (see AAY34584-Y35879) can be used in immunogenic compositions as
 CC vaccines. Vectors containing C. pneumoniae nucleotide sequences can also
 CC be used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae. (Updated on 17
 CC -OCT-2003 to standardise OS field)

XX Sequence 155 AA;

Query Match 7.0%; Score 223.5; DB 2; Length 155;
 Best Local Similarity 35.5%; Pred. No. 7.5e-14;
 Matches 50; Conservative 29; Mismatches 55; Indels 7; Gaps 2;
 QY 226 VLPPHANHQNTFGQIMAMMENAVTTIASRLCRAPTLKAIEMHFPGSPQVGRVLVKA 285
 DB 19 IFPNDIANNNTVFGGLMSLDRLVLVAERHTESVCVTAFLVDAIRFAPAYMGENLICK 78
 QY 286 AIVNNAFKHSMEVGVCEA---YRQEAETHRRHINSAPMTEFVLADDOQLLFWIRPQ 342
 DB 79 AAVNRTRTSLEVGVKVAENIYKOE---RRHTSAVTFVAVVEDNQPIPVHIVBET 134
 QY 343 GDGERRYREASARKKIRLDRK 363
 DB 135 PEKRRYNEADRRRQARLEK 155

RESULT 11

ABU26786
 ID ABU26786 standard; protein, 155 AA.

XX ABU26786;
 AC
 XX 23-OCT-2003 (revised)
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #12313.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 KW Chlamydia pneumoniae.
 XX OS
 XX WO200277183-A2.
 XX
 PD 03-OCT-2002.
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-PDB; ACA30656.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 54710; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of

the 621 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to standardise OS field)

OY 260 AHPTKATEMHERGPGVQGRILVKATVNNAFKSHMEVGYEAYRQEAETHRRH-INS 318
Db 226 TWIYTA5VDALNFHDKIKKCVITISGRMTFTSNKSMELIYVLDAPYVDSQKRYRAAS 285
OY 319 AFMTFVVLADDDQFOLLPWIRPQDGBRRRYREASAR 355
Db 286 AFFTVYSLSQEGRSLPVQVLTVEDEKRRFEKGR 322

RESULT 14
ABO34542 ID ABO34542 standard; protein; 339 AA.
XX ABO34542;
AC
XX 22-SEP-2003 (first entry)
XX
XX Region of human secreted protein encoded by cDNA sequence #169.
XX
XX Human; secreted protein; hyperproliferative disorder; blood-related disorder;
KM breast cancer; wound; reproductive disorder; lymphocytic hypoplasia;
KM haemophilia; thrombocytopaenia; immunodeficiency; multiple sclerosis;
KM Wiskott-Aldrich syndrome; autoimmune disorder; allergic asthma;
KM graft-versus-host disease; Hashimoto's thyroiditis; allergy; sepsis;
KM viral infection; bacterial infection; fungal infection; AIDS; sepsis;
KM renal disorder; kidney failure; cardiovascular disorder; cytostatic;
KM angina pectoris; cerebral ischaemia; congenital heart defect;
KM respiratory disorder; neurological disorder; Alzheimer's disease;
KM Parkinson's disease; inflammation; Crohn's disease; vulnery;
KM immunosuppressive; antibacterial; haemostatic; thrombolytic;
KM anticoagulant; neuroprotective; chymotrypsin; antiallergic;
KM antiasthmatic; virucide; fungicide; anti-HIV; nephrotoxic; anti-anginal;
KM cerebroprotective; cardiac; nootropic; antiparkinsonian;
KM antiinflammatory.
XX
XX Homo sapiens.
XX OS
XX US2003049618-A1.
XX
XX 13-MAR-2003.
XX
XX 16-MAR-2001; 2001US-00809391.
XX
XX 07-MAR-1997; 97US-0038621P.
PR 07-MAR-1997; 97US-0040162P.
PR 07-MAR-1997; 97US-0040163P.
PR 07-MAR-1997; 97US-0040333P.
PR 07-MAR-1997; 97US-0040334P.
PR 07-MAR-1997; 97US-0040336P.
PR 07-MAR-1997; 97US-0040626P.
PR 11-APR-1997; 97US-0043311P.
PR 11-APR-1997; 97US-0043312P.
PR 11-APR-1997; 97US-0043313P.
PR 11-APR-1997; 97US-0043314P.
PR 11-APR-1997; 97US-0043315P.
PR 11-APR-1997; 97US-0043568P.
PR 11-APR-1997; 97US-0043569P.
PR 11-APR-1997; 97US-0043576P.
PR 11-APR-1997; 97US-0043578P.
PR 11-APR-1997; 97US-0043580P.
PR 11-APR-1997; 97US-0043669P.
PR 11-APR-1997; 97US-0043670P.
PR 11-APR-1997; 97US-0043671P.
PR 11-APR-1997; 97US-0043672P.
PR 11-APR-1997; 97US-0043674P.
PR 23-MAY-1997; 97US-0047492P.
PR 23-MAY-1997; 97US-0047500P.
PR 23-MAY-1997; 97US-0047501P.
PR 23-MAY-1997; 97US-0047502P.
PR 23-MAY-1997; 97US-0047503P.
PR 23-MAY-1997; 97US-0047581P.
PR 23-MAY-1997; 97US-0047582P.

PR 23-MAY-1997; 97US-0047583P.
PR 23-MAY-1997; 97US-0047584P.
PR 23-MAY-1997; 97US-0047585P.
PR 23-MAY-1997; 97US-0047586P.
PR 23-MAY-1997; 97US-0047587P.
PR 23-MAY-1997; 97US-0047588P.
PR 23-MAY-1997; 97US-0047589P.
PR 23-MAY-1997; 97US-0047590P.
PR 23-MAY-1997; 97US-0047592P.
PR 23-MAY-1997; 97US-0047593P.
PR 23-MAY-1997; 97US-0047594P.
PR 23-MAY-1997; 97US-0047595P.
PR 23-MAY-1997; 97US-0047596P.
PR 23-MAY-1997; 97US-0047597P.
PR 23-MAY-1997; 97US-0047598P.
PR 23-MAY-1997; 97US-0047599P.
PR 23-MAY-1997; 97US-0047600P.
PR 23-MAY-1997; 97US-0047601P.
PR 23-MAY-1997; 97US-0047612P.
PR 23-MAY-1997; 97US-0047613P.
PR 23-MAY-1997; 97US-0047614P.
PR 23-MAY-1997; 97US-0047615P.
PR 23-MAY-1997; 97US-0047617P.
PR 23-MAY-1997; 97US-0047632P.
PR 23-MAY-1997; 97US-0047633P.
PR 06-JUN-1997; 97US-0048964P.
PR 06-JUN-1997; 97US-0048974P.
PR 13-JUN-1997; 97US-0049610P.
PR 08-JUL-1997; 97US-0051926P.
PR 16-JUL-1997; 97US-0052874P.
PR 18-AUG-1997; 97US-0055724P.
PR 22-AUG-1997; 97US-0056630P.
PR 22-AUG-1997; 97US-0056631P.
PR 22-AUG-1997; 97US-0056632P.
PR 22-AUG-1997; 97US-0056636P.
PR 22-AUG-1997; 97US-0056637P.
PR 22-AUG-1997; 97US-0056662P.
PR 22-AUG-1997; 97US-0056664P.
PR 22-AUG-1997; 97US-0056845P.
PR 22-AUG-1997; 97US-0056862P.
PR 22-AUG-1997; 97US-0056864P.
PR 22-AUG-1997; 97US-0056872P.
PR 22-AUG-1997; 97US-0056874P.
PR 22-AUG-1997; 97US-0056875P.
PR 22-AUG-1997; 97US-0056877P.
PR 22-AUG-1997; 97US-0056878P.
PR 22-AUG-1997; 97US-0056879P.
PR 22-AUG-1997; 97US-0056880P.
PR 22-AUG-1997; 97US-0056881P.
PR 22-AUG-1997; 97US-0056882P.
PR 22-AUG-1997; 97US-0056884P.
PR 22-AUG-1997; 97US-0056886P.
PR 22-AUG-1997; 97US-0056903P.
PR 22-AUG-1997; 97US-0056908P.
PR 22-AUG-1997; 97US-0056909P.
PR 22-AUG-1997; 97US-0056910P.
PR 22-AUG-1997; 97US-0056911P.
PR 05-SEP-1997; 97US-0057650P.
PR 05-SEP-1997; 97US-0057669P.
PR 05-SEP-1997; 97US-0057761P.
PR 12-SEP-1997; 97US-0058785P.
PR 09-OCT-1997; 97US-0061660P.
PR 06-MAR-1998; 98WO-US0004493.
PR 08-SEP-1998; 98US-00149476.
PR 17-MAR-2000; 2000US-0190068P.

XX (RUBE/) RUBEN S. M.
PA (ROSE/) ROSEN C. A.
PA (SOPP/) SOPPET D. R.
PA (CART/) CARTER K. C.
PA (BEDN/) BEDNARIK D. P.
PA (ENDR/) ENDRESS G. A.
PA (YUGG/) YU G.
PA (NIJU/) NI J.
PA (FENG/) FENG P. P.
PA (YOUN/) YOUNG P. E.
PA (GREE/) GREENE J. M.
PA (FERR/) FERRIE A. M.
PA (DUAN/) DUAN D. R.
PA (HUJU/) HU J.
PA (FLOR/) FLORENCE K. A.
PA (OLSE/) OLSEN H. S.
PA (FISC/) FISCHER C. L.
PA (EBNE/) EBNER R.
PA (BREW/) BREWER L. A.
PA (MOOR/) MOORE P. A.
PA (SHIY/) SHI Y.
PA (LAFL/) LAFLAUR D. W.
PA (LIYY/) LI Y.
PA (ZENG/) ZENG Z.
PA (KYAW/) KYAW H.
XX
PI Ruben SM, Rosen CA, Soppet DR, Carter KC, Bednarik DP;
PI Endress GA, Yu G, Ni J, Feng P, Young PE, Greene JM, Ferrie AM;
PI Duan DR, Hu J, Florence KA, Olsen HS, Fischer CL, Ebner R,
PI Brewer LA, Moore PA, Shi Y, Laflaur DW, Li Y, Zeng Z, Kyaw H;
XX
XX WPI: 2003-521800/49.
DR N-PSDB; ACD82809.
XX
XX New genes and its encoded prostate cancer antigen proteins, useful for
PT preventing, treating, ameliorating or diagnosing e.g. prostate cancers,
PT thymic hypoplasia, multiple sclerosis, AIDS, angina pectoris or cerebral
PT ischemia.
XX
PS Claim 3; SEQ ID NO 488; 260pp; English.
XX
CC The present invention relates to the isolation of novel human secreted
CC proteins and the polynucleotide sequences encoding them. The invention
CC also discloses vectors, host cells, antibodies, and recombinant methods
CC for producing human secreted proteins. The polypeptide and polynucleotide
CC sequences for the secreted proteins are useful for preventing, treating,
CC ameliorating or diagnosing medical conditions such as hyperproliferative
CC disorders (e.g. leukemia or breast cancers), wounds, reproductive
CC disorders, blood-related disorders (e.g. haemophilia or
CC thrombocytopaenia), immunodeficiencies (e.g. Wiskott-Aldrich syndrome or
CC thymic hypoplasia), autoimmune disorders (e.g. graft-versus-host disease,
CC multiple sclerosis or Hashimoto's thyroiditis), allergies (e.g. asthma),
CC viral or bacterial or fungal infections (e.g. AIDS or sepsis), renal
CC disorders (e.g. kidney failure), cardiovascular disorders (e.g. angina
CC pectoris, cerebral ischaemia or congenital heart defects), respiratory
CC disorders, neurological disorders (e.g. Alzheimer's disease or
CC Parkinson's disease), and inflammations (e.g. Crohn's disease). The
CC polynucleotide or polypeptide may also be used as vaccine adjuvants.
CC AB034374-AB034815 represent human secreted proteins or their fragments.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site at seqdata.uspto.gov/pspsidentry.html
XX
SQ Sequence 339 AA:
Query Match 6.9%; Score 220; DB 6; Length 339;
Best Local Similarity 22.0%; Pred. No. 6.2e-13;
Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

QY 38 GGGYRNTEVOMGQLVPCHTNNGELSYGQLKWDITTCASERRAG-----CCCVT 91
DB 3 GPDVETPSAIQICIKMRPDDANVAGNVHGTLIKMIEAGATISTKHCNSONGERCVAAL 62
QY 92 ASMDIYFEHTISVGQYVNIKAKVNFNSMEVGIOVASEDCSCQKNNVCALATFVA 151
DB 63 ARVERTDFLSPMCGEVAHSAEITVTSKDSVEQVNVMSENILTGAKKLTNRKATLVVP 122
QY 152 RREITTKYKQIIP-----RTEEF-----KMHSAVERRRRLVYADITIKOLLANCAIQGD 203
DB 123 LSLKNVDKVLLEVPVVSXKXOEGRKRYEAKQLEMETKWRNGDIVQPLN----- 175
QY 204 LESRDCSRMVAPEKT-----RVESVELVPPHANQGTFFGQIWMAMENVTATIASRLCR 259
DB 176 -----PEPVTYSQSSLIHLVGEDCTIAGFVHGCVTKMLMDVAGIYAAHRCK 225
QY 260 AHPITKAIEHFHRRGSPQVSDRLVTKAIVNNAFGRSMVEGVCVAYRQEAETHRRH-INS 318
DB 226 TNIVTASVDALINFDKIRKGCVITISGRMTFTSNKSMIEIVLVADPVDVSSQKRYRAAS 285
QY 319 AFMTFVVLDDDDPQILPMTRPQDGERRYRREASAR 355
DB 286 AFTYVYSLSOEGRSLPVPQLVPETEDKKRFEKGR 322
RESULT 15
IDA55399
ID ADA55399 standard; protein; 370 AA.
XX
AC ADA55399;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 2967.
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
XX 21-MAR-2002; 2002EP-00006586.
XX
XX 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
DR WPI: 2003-395539/38.
DR N-PSDB; ADA553760.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 14; SEQ ID NO 2967; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA552433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 370 AA:
Query Match 6.9%; Score 219; DB 6; Length 370;
Best Local Similarity 22.0%; Pred. No. 9.1e-13;

Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

```
QY 38 GEGYRNPTVEVQMSQVLPECHTNOBGLSVGLKMTIDTACLSARHAG-----CPCVT 91
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 35 GPDVETPSAIOICRIMRPDANVAGNVHGGTILKMIKGAGALISTRHCNSONGERCVAAL 94
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 92 ASMDIYFEHTISVQVYVNIKAKVNRAFNSSMVEGIVASEDLCEKQMNVCALATFVA 151
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 95 ARVERTDLSIPMCIGEVAVHSAEITYTSKHSVEYQVYVNMSENILTGAKKLTNKAFLMYVP 154
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 152 RREITKVKLKQITP-----RTEEEKMEHSVAERRRMELVY--ADTIKDLILANCAIQGD 203
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 155 LSLKXVDKYLEVPPVYVSGROEBOEBGRKRYEAKLERMETKWRNGDIVQPVLN----- 207
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 204 LESRDCSMVPAEKT---RVESVELVLPRHMHQGNTPFGQIMAMMENVATIASRLCR 259
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 208 -----PEPNTVSYSSQSLIHVLGSPDCTLHGFVHGVTMKLMDDEVAGIVAAHCK 257
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 260 AHPYLKALEMFHFRGSPGVGDRLVLKATVNNAFKHSMEVGVCVEAYROEAETHRRH-INS 318
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 258 TNIVTASVDALNFDHXIKRKGCVITISGRMTFTSNKSMELIVLVDADPVVDSQKRYRAAS 317
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 319 APMTFVVLADDDQQLLPWIRPQPGDGERRYREASAR 355
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 318 AFTYVSLSQEGRSLPVPQLVPEDEDEKKRFEKGGR 354
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

Search completed: August 12, 2004, 15:44:04
Job time : 61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2004, 23:35:26 ; Search time 883 Seconds

(without alignments)
10330.500 Million cell updates/sec

Title: US-10-055-624B-1

Perfect score: 1857
Sequence: 1 gtggaattgccttcacat.....cctgaacctatcatgagc 1857

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3228839 seqs, 245606551 residues

Total number of hits satisfying chosen parameters: 6457678

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1857	100.0	1857	16	US-10-055-624B-1
2	1839	99.0	6359	17	US-10-055-624B-1
3	1633.6	88.0	1818	16	US-10-055-624B-3
4	1230	26.2	2699	16	US-10-055-624B-5
5	471.8	25.4	813	13	US-09-925-298-181
6	471.8	25.4	813	15	US-10-102-806-181
7	403.4	21.7	2034	13	US-10-336-472-49
8	283.6	15.3	2460	17	US-10-416-314-89
9	143.2	7.7	2242	16	US-10-108-260A-1210
10	126.6	6.8	390	9	US-09-867-701-2605
11	126.6	6.8	573	10	US-09-814-353-21183
12	124.6	6.7	572	10	US-09-814-353-17043
13	124	6.7	543	10	US-09-918-995-21270
14	117.8	6.3	452	9	US-09-864-761-2374

15	115.2	6.2	594	13	US-10-027-632-191598	Sequence 191598,
16	115.2	6.2	594	13	US-10-027-632-191599	Sequence 191599,
17	115.2	6.2	594	13	US-10-027-632-191600	Sequence 191600,
18	115.2	6.2	594	16	US-10-027-632-191598	Sequence 191598,
19	115.2	6.2	594	16	US-10-027-632-191599	Sequence 191599,
20	115.2	6.2	594	16	US-10-027-632-191600	Sequence 191600,
21	114.2	6.1	629	10	US-09-814-353-10659	Sequence 10659, A
22	114.2	6.1	629	10	US-10-027-632-231207	Sequence 231207,
23	112	6.0	500	13	US-10-027-632-231208	Sequence 231208,
24	112	6.0	500	13	US-10-027-632-231209	Sequence 231209,
25	112	6.0	500	13	US-10-027-632-231209	Sequence 231209,
26	112	6.0	500	16	US-10-027-632-231207	Sequence 231207,
27	112	6.0	500	16	US-10-027-632-231208	Sequence 231208,
28	112	6.0	500	16	US-10-027-632-231209	Sequence 231209,
29	101	5.4	492	9	US-09-864-761-1743	Sequence 1743, Ap
30	100.6	5.4	1204	17	US-10-115-635-272	Sequence 272, App
31	100.6	5.4	2252	16	US-10-104-047-140	Sequence 140, App
32	100.6	5.4	2255	16	US-10-104-047-881	Sequence 881, App
33	100.6	5.4	137870	13	US-10-351-951-1	Sequence 1, Appl
34	98.8	5.3	282	10	US-09-814-353-21609	Sequence 21609, A
35	98.8	5.3	385	10	US-09-814-353-18091	Sequence 18091, A
36	97.6	5.3	263	10	US-09-814-353-5420	Sequence 5420, Ap
37	97.6	5.3	263	10	US-09-814-353-11707	Sequence 11707, A
38	95.4	5.1	47243	13	US-10-087-192-1264	Sequence 1264, App
39	94.6	5.1	2950	16	US-10-108-260A-165	Sequence 165, App
40	89.8	4.8	844	13	US-10-027-632-149777	Sequence 149777,
41	89	4.8	844	13	US-10-027-632-149777	Sequence 149777,
42	89	4.8	844	16	US-10-027-632-149777	Sequence 149777,
43	88.2	4.7	905	16	US-10-120-988-311	Sequence 311, App
44	79	4.3	621	15	US-10-156-761-5013	Sequence 5013, Ap
45	79	4.3	9025608	15	US-10-156-761-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-055-624B-1
; Sequence 1, Appl, Application US/10055624B
; Publication No. US2003022038A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Sean H
; APPLICANT: Chui, Clarissa
; APPLICANT: Goddard, Audrey D
; APPLICANT: Grimaldi, J. Christopher
; TITLE OF INVENTION: BPIR COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 9800081-0066
; CURRENT APPLICATION NUMBER: US/10/055, 624B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/263,362
; PRIOR FILING DATE: 2002-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1857
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-055-624B-1

Query Match 100.0%; Score 1857; DB 16; Length 1857;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1857; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGAATTGCCCTTCAATGATCCAGAAATGCGAAATCAGTCGGAGCGGCTTGCC 60
DB 1 GTGGAATTGCCCTTCAATGATCCAGAAATGCGAAATCAGTCGGAGCGGCTTGCC 60
QY TCTGTGTTCTCCACCGCATCCCGGAAGTCACTTACGTGCGGGAACGACAGTCC 120
DB TCTGTGTTCTCCACCGCATCCCGGAATTCAGCTTACGTGCGGGAACGACAGTCC 120
QY ATGGCAGACGGCGGAGATCCCGGAATCCCGGAGGTGAGATGACCGACTGTGCTG 180
DB ATGGCAGACGGCGGAGATCCCGGAATCCCGGAGGTGAGATGACCGACTGTGCTG 180

Db 121 ATGGAGACGGGAGGATACCGAACCACGAGGTGAGATGAGCCAGTGTGCTG 180
QY 181 CCGTGGCCACACCAACCAAGTGTGAGCTGCGGCGAGCTCTCAAGTGGATTGAC 240
Db 181 CCGTGGCCACACCAACCAAGTGTGAGCTGCGGCGAGCTCTCAAGTGGATTGAC 240
QY 241 ACCAGGCTGTGCTGCGGAGAGGAGCGCTGCGCTGCTGCTGCTGCTGCTGCTG 300
Db 241 ACCAGGCTGTGCTGCGGAGAGGAGCGCTGCGCTGCTGCTGCTGCTGCTGCTG 300
QY 301 GATGACATCTATTTTGTGACACCAATTAGTGTGACCAAGTGTGATATCAAGGCAAG 360
Db 301 GATGACATCTATTTTGTGACACCAATTAGTGTGACCAAGTGTGATATCAAGGCAAG 360
QY 361 GTGAACCGGCTTCACTCCAGATGAGGTGGGCACTGAGTGGCTCCGAGAGACTG 420
Db 361 GTGAACCGGCTTCACTCCAGATGAGGTGGGCACTGAGTGGCTCCGAGAGACTG 420
QY 421 TGCTCTGAGAGAGAGTGTGCAAGGCTTGGGCGACCTTGTGCTGCTGCTGCTG 480
Db 421 TGCTCTGAGAGAGAGTGTGCAAGGCTTGGGCGACCTTGTGCTGCTGCTGCTG 480
QY 481 ATCAACCAAGTGTGAGTGTGAGCAATCAAGCGGAGAGAGAGAGAGAGAGAGAG 540
Db 481 ATCAACCAAGTGTGAGTGTGAGCAATCAAGCGGAGAGAGAGAGAGAGAGAGAG 540
QY 541 AGTGTGGGCTGTGAGGCGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 541 AGTGTGGGCTGTGAGGCGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 600
QY 601 CTGGCCAACTGCGGCACTCAGGCGGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 660
Db 601 CTGGCCAACTGCGGCACTCAGGCGGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 GCTGAG 720
Db 661 GCTGAG 720
QY 721 GGCAACACCTTTGGGAGGCGAGATCAATGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 721 GGCAACACCTTTGGGAGGCGAGATCAATGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 AGCGGCTGTGCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db 781 AGCGGCTGTGCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 841 CCGTCCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db 841 CCGTCCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 AGCAGTGTGAG 960
Db 901 AGCAGTGTGAG 960
QY 961 CACATCAACAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 961 CACATCAACAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1021 CTGCGCTGTGAG 1080
Db 1021 CTGCGCTGTGAG 1080
QY 1081 AGAAG 1140
Db 1081 AGAAG 1140
QY 1141 CTCTCCGCTGTGAG 1200
Db 1141 CTCTCCGCTGTGAG 1200
QY 1201 TTGAAGATGCTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1201 TTGAAGATGCTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

QY 1261 CTGTACACTGTGAG 1320
Db 1261 CTGTACACTGTGAG 1320
QY 1321 GATGAG 1380
Db 1321 GATGAG 1380
QY 1381 CACTACCGAG 1440
Db 1381 CACTACCGAG 1440
QY 1441 ACCAGGCTGCTGTGAG 1500
Db 1441 ACCAGGCTGCTGTGAG 1500
QY 1501 CGAAG 1560
Db 1501 CGAAG 1560
QY 1561 AOCACCGAG 1620
Db 1561 AOCACCGAG 1620
QY 1621 TGCGCGAG 1680
Db 1621 TGCGCGAG 1680
QY 1681 GTCTCGGAG 1740
Db 1681 GTCTCGGAG 1740
QY 1741 GACGCTGTGAG 1800
Db 1741 GACGCTGTGAG 1800
QY 1801 GCCAAATCAACAG 1857
Db 1801 GCCAAATCAACAG 1857

RESULT 2

US-10-717-597-301
; Sequence 301, Application US/10717597
; Publication No. US20040110221A1
; GENERAL INFORMATION:
; APPLICANT: Burczynski, Michael E.
; APPLICANT: Wyeth
; APPLICANT: Twine, Natalie C.
; APPLICANT: Dornier, Andrew J.
; APPLICANT: Trepicchio, William L.
; APPLICANT: Stonim, Donna K.
; APPLICANT: Stover, Jennifer A.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND OTHER SOLID TUMORS
; FILE REFERENCE: AM101080L
; CURRENT APPLICATION NUMBER: US/10/717,597
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US 60/459,782
; PRIOR FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: US 60/427,982
; NUMBER OF SEQ ID NOS: 4904
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 301
; LENGTH: 6359
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-717-597-301

Query Match 99.0%; Score 1839; DB 17; Length 6359;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1839; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 ATGATCCAGAAATGTCGGAATCACTTCGACGCGGCTTGGCTCTGTGTCTCCAAACCG 78
DB 72 ATGATCCAGAAATGTCGGAATCACTTCGACGCGGCTTGGCTCTGTGTCTCCAAACCG 131
QY 79 ACATCCCGGAAGTCAGCTTAACTGTCGCGGGAACGACATGTCATGTCGACGCGGGA 138
DB 132 ACATCCCGGAAGTCAGCTTAACTGTCGCGGGAACGACATGTCATGTCGACGCGGGA 191
QY 139 TACCGGAATCCCGACGAGTGCAGATGAGCCAGCTGGTGGCTTCGACCAACCA 198
DB 192 TACCGGAATCCCGACGAGTGCAGATGAGCCAGCTGGTGGCTTCGACCAACCA 251
QY 199 CGTGTGAGCTGAGCGTTCGCGGACCTGCTCAAGTGAATGACACACGCGTTCGCTCC 258
DB 252 CGTGTGAGCTGAGCGTTCGCGGACCTGCTCAAGTGAATGACACACGCGTTCGCTCC 311
QY 259 GCGGAGAGGACGCTGGCTGCTGTCACAGCTTCCATGATGACATCTATTGAG 318
DB 312 GCGGAGAGGACGCTGGCTGCTGTCACAGCTTCCATGATGACATCTATTGAG 371
QY 319 CACACATTAGTGTGACAAAGTGTGATATCAAGGCCAAAGTGAACCGGCGCTTCAAC 378
DB 372 CACACATTAGTGTGACAAAGTGTGATATCAAGGCCAAAGTGAACCGGCGCTTCAAC 431
QY 379 TCCAGATGAGAGTGGGATCAGTGGCTTCGAGAGACCTGCTCTGAGAAACAGTGG 438
DB 432 TCCAGATGAGAGTGGGATCAGTGGCTTCGAGAGACCTGCTCTGAGAAACAGTGG 491
QY 439 AATGTGTGCAAGGCTTGGCCACTTGTGTGCGCCGAGAGATCAACCAAGTGAAGCTG 498
DB 492 AATGTGTGCAAGGCTTGGCCACTTGTGTGCGCCGAGAGATCAACCAAGTGAAGCTG 551
QY 499 AAGCAGATCAAGCGCGGAGACAGAAAGAGATGAGACAGTGTGGCGCTGAGCGC 558
DB 552 AAGCAGATCAAGCGCGGAGACAGAAAGAGATGAGACAGTGTGGCGCTGAGCGC 611
QY 559 CGGCGCATGCGGCTTGTCTATGACAGACACCATCAAGACCTCTGCGCAACCTGCGCAT 618
DB 612 CGGCGCATGCGGCTTGTCTATGACAGACACCATCAAGACCTCTGCGCAACCTGCGCAT 671
QY 619 CAGGCGCATGTCGAGACAGACACTGTAGCCGATGTCGCGCTGAGAAACCGCTGTG 678
DB 672 CAGGCGCATGTCGAGACAGACACTGTAGCCGATGTCGCGCTGAGAAACCGCTGTG 731
QY 679 GAGAGTGTGAGAGTGTGCTGCTGCTGCGCCGACATGACAGAGGCAACCTTGGGAGC 738
DB 732 GAGAGTGTGAGAGTGTGCTGCTGCTGCGCCGACATGACAGAGGCAACCTTGGGAGC 791
QY 739 CAGATCATGCTGAGATGAGAAATGTGACCACTTCAGCGACGCGCTCTGCGCTGCGC 798
DB 792 CAGATCATGCTGAGATGAGAAATGTGACCACTTCAGCGACGCGCTCTGCGCTGCGC 851
QY 799 CACCTTACGCTGAAGGCTATTTAAATGTTCCATTCGAGGCGCTTCGAGAGTGCAGC 858
DB 852 CACCTTACGCTGAAGGCTATTTAAATGTTCCATTCGAGGCGCTTCGAGAGTGCAGC 911
QY 859 CGTCTGCTGCTCAAGGCTATGTAACAATGCTTCAAACTACAGAGAGTGGGCTG 918
DB 912 CGTCTGCTGCTCAAGGCTATGTAACAATGCTTCAAACTACAGAGAGTGGGCTG 971
QY 919 TGGGTGAGAGGCTATTCGACAGAGGCTGAGACCAACGCGCGCTCATCAACAGTGCCTT 978
DB 972 TGGGTGAGAGGCTATTCGACAGAGGCTGAGACCAACGCGCGCTCATCAACAGTGCCTT 1031
QY 979 ATGACCTTTGTGTCTGAGACGAGATGACCAAGCTTCGCTGCTGCTGAGTTCGCGC 1038
DB 1032 ATGACCTTTGTGTCTGAGACGAGATGACCAAGCTTCGCTGCTGCTGAGTTCGCGC 1091
QY 1039 CAGCGCGGAGATGTCGAGCGCGCTTACGAGAGGCGAGTGCAGAAAGAAATGTCGCGCT 1098
DB 1092 CAGCGCGGAGATGTCGAGCGCGCTTACGAGAGGCGAGTGCAGAAAGAAATGTCGCGCT 1151

QY 1099 GACAGGAATGATCTGCTGCTGTAAGACAGAGAGTGCCTCTTCGCTCCCTGGGAC 1158
DB 1152 GACAGGAATGATCTGCTGCTGTAAGACAGAGAGTGCCTCTTCGCTCCCTGGGAC 1211
QY 1159 CCTAGCAACAGGATGATCTGAGCTCAATTAAGTCTCCCTTGAAGATGCTTGGCC 1218
DB 1212 CCTAGCAACAGGATGATCTGAGCTCAATTAAGTCTCCCTTGAAGATGCTTGGCC 1271
QY 1219 AAGACAACTGGTGTGCTGCTGAGATGATCAGTTCGCTGTAACCTGAGAT 1278
DB 1272 AAGACAACTGGTGTGCTGCTGAGATGATCAGTTCGCTGTAACCTGAGAT 1331
QY 1279 GACAGTCTCTCTTCCATGAGATGATGATGATGATGATGATGATGATGATGATGAT 1338
DB 1332 GACAGTCTCTCTTCCATGAGATGATGATGATGATGATGATGATGATGATGATGAT 1391
QY 1339 CTGCTGCTCTGAGACCTGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 1398
DB 1392 CTGCTGCTCTGAGACCTGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAG 1451
QY 1399 CTAGTGCAGAGGATGACGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1458
DB 1452 CTAGTGCAGAGGATGACGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1511
QY 1459 GGTCAACAAAGCCCGAGACTTGTGATCTGAGCTTCGAGGAGGAGGAGGAGGAGGAG 1518
DB 1512 GGTCAACAAAGCCCGAGACTTGTGATCTGAGCTTCGAGGAGGAGGAGGAGGAGGAG 1571
QY 1519 GGGAGCCCTTATGATCATGCGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTG 1578
DB 1572 GGGAGCCCTTATGATCATGCGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTG 1631
QY 1579 GAGTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1638
DB 1632 GAGTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1691
QY 1639 CTGACCAAGTGTGCTGAGGATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1698
DB 1692 CTGACCAAGTGTGCTGAGGATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1751
QY 1699 TATTCCTGGGAGCTGCAATCCAGTGAAGGCTGAGGCTGAGGCTGAGGCTGAGGAG 1758
DB 1752 TATTCCTGGGAGCTGCAATCCAGTGAAGGCTGAGGCTGAGGCTGAGGCTGAGGAG 1811
QY 1759 CTAGTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1818
DB 1812 CTAGTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1871
QY 1819 TTTGATACCTTCAAGACACTGAAACCTTATCATGAGC 1857
DB 1872 TTTGATACCTTCAAGACACTGAAACCTTATCATGAGC 1910

RESULT 3
US-10-055-624B-3
; Sequence 3, Application US/10055624B
; Publication No. US20030220238A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Sean H
; APPLICANT: Choi, Clarissa
; APPLICANT: Goddard, Audrey D
; APPLICANT: Grimaldi, J. Christopher
; TITLE OF INVENTION: BPT COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 980081-0066
; CURRENT APPLICATION NUMBER: US/10/055,624B
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/263,362
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 1818
; TYPE: DNA

OTHER INFORMATION: n equals a,t,g, or c
US-09-925-298-181

Query Match 25.4%; Score 471.8; DB 13; Length 813;
Best Local Similarity 96.4%; Pred. No. 3e-128;
Matches 487; Conservative 5; Mismatches 12; Indels 1; Gaps 1;

```
QY 1162 AGCAACCGAGGTACTGCTGATCAATATACGCTCTCTGAAATGCTTGGCCAA 1221
DB 9 AGAGACGAGGTGTACTGATCAATATACGCTCTCTGAAATGCTTGGCCAA 68
QY 1222 GACAACTGGGTGTCTCTCTGAGATCACTAGTCCGCTGTACATCTGGAGATAC 1281
DB 69 GACAACTGGGTGTCTCTCTGAGATCACTAGTCCGCTGTACATCTGGAGATAC 128
QY 1282 AAGTTCTCTCTCTCTCAATGAGATGCTGTGATGATGACGCCAGGCTTCTG 1341
DB 129 AAGTTCTCTCTCTCTCAATGAGATGCTGTGATGATGACGCCAGGCTTCTG 188
QY 1342 CTGCTCTGAGCTGTGCTGAGAGCGAGATGAGCACTACCGAGAGGTGAGCTA 1401
DB 189 CTGCTCTGAGCTGTGCTGAGAGCGAGATGAGCACTACCGAGAGGTGAGCTA 248
QY 1402 GTGAGAGAGGTAGAGCA-GGAGAGAGCGATCTACAGCTGACGAGCTTGGAG 1460
DB 249 GTGAGAGAGGTAGAGCA-GGAGAGAGCGATCTACAGCTGACGAGCTTGGAG 308
QY 1461 TCACCAAAAGCCCGAGACTTCTGTATCTGCTGAGAGCGAGAGCTTGTGACAT 1520
DB 309 TCACCAAAAGCCCGAGACTTCTGTATCTGCTGAGAGCGAGAGCTTGTGACAT 368
QY 1521 GGAACCCATGTCATCGGCTGAGGTGCTGACGCTGCGCAGAGCGAGAGCGCAGA 1580
DB 369 GGAACCCATGTCATCGGCTGAGGTGCTGACGCTGCGCAGAGCGAGAGCGCAGA 428
QY 1581 GTACAGAGCGGAGAGAGAGCTTCTGCTGAGGCTTCTGCTGAGAGCGAGAGCACT 1640
DB 429 GTACAGAGCGGAGAGAGAGCTTCTGCTGAGGCTTCTGCTGAGAGCGAGAGCACT 488
QY 1641 GACCAAGTGTCTGCTGAGGCTTGGGCTC 1665
DB 489 GACCAAGTGTCTGCTGAGGCTTGGGCTC 513
```

RESULT 6

US-10-102-806-181
Sequence 181, Application US/10102806
Publication No. US20030054421A1

GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA103PCT
CURRENT APPLICATION NUMBER: US/10/102,806
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: 09/925,298
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05881
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 846
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 181

LENGTH: 813
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (266)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (723)
OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc feature
LOCATION: (726)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (738)
OTHER INFORMATION: n equals a,t,g, or c
US-10-102-806-181

Query Match 25.4%; Score 471.8; DB 15; Length 813;
Best Local Similarity 96.4%; Pred. No. 3e-128;
Matches 487; Conservative 5; Mismatches 12; Indels 1; Gaps 1;

```
QY 1162 AGCAACCGAGGTACTGCTGATCAATATACGCTCTCTGAAATGCTTGGCCAA 1221
DB 9 AGAGACGAGGTGTACTGATCAATATACGCTCTCTGAAATGCTTGGCCAA 68
QY 1222 GACAACTGGGTGTCTCTCTGAGATCACTAGTCCGCTGTACATCTGGAGATAC 1281
DB 69 GACAACTGGGTGTCTCTCTGAGATCACTAGTCCGCTGTACATCTGGAGATAC 128
QY 1282 AAGTTCTCTCTCTCTCAATGAGATGCTGTGATGATGACGCCAGGCTTCTG 1341
DB 129 AAGTTCTCTCTCTCTCAATGAGATGCTGTGATGATGACGCCAGGCTTCTG 188
QY 1342 CTGCTCTGAGCTGTGCTGAGAGCGAGATGAGCACTACCGAGAGGTGAGCTA 1401
DB 189 CTGCTCTGAGCTGTGCTGAGAGCGAGATGAGCACTACCGAGAGGTGAGCTA 248
QY 1402 GTGAGAGAGGTAGAGCA-GGAGAGAGCGATCTACAGCTGACGAGCTTGGAG 1460
DB 249 GTGAGAGAGGTAGAGCA-GGAGAGAGCGATCTACAGCTGACGAGCTTGGAG 308
QY 1461 TCACCAAAAGCCCGAGACTTCTGTATCTGCTGAGAGCGAGAGCTTGTGACAT 1520
DB 309 TCACCAAAAGCCCGAGACTTCTGTATCTGCTGAGAGCGAGAGCTTGTGACAT 368
QY 1521 GGAACCCATGTCATCGGCTGAGGTGCTGACGCTGCGCAGAGCGAGAGCGCAGA 1580
DB 369 GGAACCCATGTCATCGGCTGAGGTGCTGACGCTGCGCAGAGCGAGAGCGCAGA 428
QY 1581 GTACAGAGCGGAGAGAGAGCTTCTGCTGAGGCTTCTGCTGAGAGCGAGAGCACT 1640
DB 429 GTACAGAGCGGAGAGAGAGCTTCTGCTGAGGCTTCTGCTGAGAGCGAGAGCACT 488
QY 1641 GACCAAGTGTCTGCTGAGGCTTGGGCTC 1665
DB 489 GACCAAGTGTCTGCTGAGGCTTGGGCTC 513
```

RESULT 7

US-10-336-472-49
Sequence 49, Application US/10336472
Publication No. US20040043929A1

GENERAL INFORMATION:
APPLICANT: Anderson, David W.
APPLICANT: Ballinger, Robert A.
APPLICANT: Baumgartner, Jason C.
APPLICANT: Burgess, Catherine E.
APPLICANT: Caeman, Stacie J.
APPLICANT: Chant, John S.
APPLICANT: Berghs, Constance
APPLICANT: Gangoli, Bsha A.
APPLICANT: Edinger, Shlomit R.
APPLICANT: Ellerman, Karen
APPLICANT: Futrak, Katarzyna
APPLICANT: Gerlach, Valerie
APPLICANT: Gilbert, Jennifer A.
APPLICANT: Gunther, Erik
APPLICANT: Gorman, Linda
APPLICANT: Guo, Xiaojia Sasha
APPLICANT: Ji, Weizhen
APPLICANT: Li, Li
APPLICANT: Liu, Xiaohong

APPLICANT: Miller, Charles E.
APPLICANT: Padigaru, Muralidhara
APPLICANT: Rastellari, Luca
APPLICANT: MacDougall, John R.
APPLICANT: Mishra, Vishnu
APPLICANT: Pena, Carol E.A.
APPLICANT: Spaderna, Steven K.
APPLICANT: Shimkels, Richard A.
APPLICANT: Smithson, Glennda
APPLICANT: Spytek, Kimberly A.
APPLICANT: Stone, David J.
APPLICANT: Shenoy, Suresh G.
APPLICANT: Ort, Tatiana
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Tcherenev, Velizar T.
APPLICANT: Verne, Corine A.M.
APPLICANT: Wolenc, Adam R.
APPLICANT: Zernusen, Bryan D.
APPLICANT: Zhong, Mei
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-533C
CURRENT APPLICATION NUMBER: US/10/336,472
CURRENT FILING DATE: 2003-01-03
PRIOR APPLICATION NUMBER: 09/746,491
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 10/005,041
PRIOR FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 10/023,681
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: 10/024,212
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: 10/055,569
PRIOR FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: 10/080,334
PRIOR FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: 10/092,900
PRIOR FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: 10/136,826
PRIOR FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: 10/236,417
PRIOR FILING DATE: 2002-09-06
PRIOR APPLICATION NUMBER: 60/345,092
PRIOR FILING DATE: 2002-01-04
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 230
SOFTWARE: Cureseqlist version 0.1
SEQ ID NO 49
LENGTH: 2034
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (61)..(1725)
US-10-336-472-49

Query Match 21.7%; Score 403.4; DB 13; Length 2034;

Best Local Similarity 56.0%; Pred. No. 5,3e-108;

Matches 849; Conservative 0; Mismatches 651; Indels 15; Gaps 4;

QY 115 AGTGCATGCGAGACGGGAGATACCGGAAACCCACGGAGGTGCAGATGAGCCAGCTG 174
DB 43 AGGCTCTCGCCCTGCGCATGAGAGCGGCCGCGCCGCGCGAGGTGTCATGAGCCAGACC 102
QY 175 GTGCTCCCTGCGACACCAACCAACGCTGAGCTGAGCTCGGCGAGCTGCTCAAGTG 234
DB 103 ATCAAGCCGCGCAGCGCATCTGCGCGGCGGAGCTGAGCGCGGCGAGCTGCTCAAGTG 162
QY 235 ATTGACACCAACGCTGCTGCTGCTCGGAGAGGACGCTGCTGCTGCTGCTGCTGCTGCT 294
DB 163 ATGGAACCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 222

QY 295 TCATGATGATCATCTATTTTGTAGACACCACTTATGTTGTGACAAAGTGTGATATCAAG 354
DB 223 TCAGTGTATGATCATATACAGTTTGTAGAGACAGCTAGATGTGACAAAGTTATATACCATCAA 282
QY 355 GCCAAGGTGAACCGGCTTTCATCTCCAGCATGTGAGGTGGCATTCAGGTGGCTCGGAG 414
DB 283 GCAAAAGTATCTAAGATTCAGCAAGAGATGAGATCATGATCAAGATCATGATGATGATG 342
QY 415 GACCTGTCTCTGAGAGACAGTGAATGTGTGCAAGGCTTGGCCACTTGTGGCCCGC 474
DB 343 GATATGCTCACTGACATTTGAGAAAGCTTTTATGTTGCTTTCTCCACTTTTGTAGCCAAA 402
QY 475 CGA---GAGATCAACAAAGTGAAGCTGAAGCATCAAGCCGCGACAGAGAGAGAGAG 531
DB 403 CCAATTGGAAGAAAGAAAGATTCATTTAAACCACTCAACATTTCACTGAACCAAGATCAT 462
QY 532 ATGAGACAGATGTGGGGGCTGAGCCGCGGACATGGGCTTGTCTATGACACATATC 591
DB 463 GTGAAACATATATCTGCTGCTGAGAGAGAAAGTTGATTAACAACATGAAATATCTTT 522
QY 592 AAGGACTCTGCGCACTGCGCATTCAGGCGCATCTGAGAGACAGACTGATGCGC 651
DB 523 AACATTTTATGAAAGAAAGTGAATTTGATGATCTCATTTTGTATGAAGAGAAAGA 582
QY 652 ATGCTGCGGCTGAAGAAAGCCGCTGTGAGAGTGTGAGCTGATCTGCTCCCAAGCC 711
DB 583 GCGGTTTCCAAAGAGGACCTCCGTTCAAGAGCATTAATGATGATGATGATGATGATGATG 642
QY 712 AATCAACAGGCAACCTTTGGGGGCAATATATGCTGATGAGATGTGGCAAC 771
DB 643 AACCATCAAGAAATATTTGTTGGCAATTAATGCGTGAAGAGAGACAGTGTACT 702
QY 772 ATTGACAGCCGCGCTCTGCGCTGCGCCCACTTACGCTGAAGGCACTGAATGTGCAC 831
DB 703 ATTTCGACAGCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 762
QY 832 TTCGAGGCGCGTCCAGAGTCCGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891
DB 763 TTCGAGGCGCGTCCAGAGTCCGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 822
QY 892 TTCAACATATGATGAGGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 945
DB 823 TTTCAGACCTGTGTGAGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 882
QY 946 GAGACCCAGCGGCGCCATCAACAGTGCCTTATATGCTTGTGAGTGTGAGTGTGAGTGTGAGT 1005
DB 883 GAGGCGCGAGGCGCTGATCAATCAACAGTGTGCTTCTCTATTAACAATTTGCTGATGATGAG 942
QY 1006 GACCAAGCCCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1065
DB 943 GAAATCTCATCAAGTTTCCCAAGATCAACCATTTCAAGGATGATTTCAAGCGCTAT 1002
QY 1066 CGAGAGCCAGTCCGAGAAAGATTCGCTGAGACAGAAAGTATCTGCTGCTGCTGCTGCTGCT 1125
DB 1003 CGGAGAGCTATTCAGCGAAGCGAATTCGCTGAGGCAAAATATGTTATTTCCCAAAA 1062
QY 1126 CAGACAGAGTGCCTCTCCCTGCGGACCTTATGACCAACAGGCTGCTGAGCTATC 1185
DB 1063 GA---AGAGTTTCACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119
QY 1186 AATACGCTCTCTCTTGAAGATGCTTGTGCGCAAGACAACTGGTGTCTGCTGCTGCTGCTG 1245
DB 1120 AGCAATGTGAGGCGCTCAAAAATACTGCGAGCCAAAAGGGGTTGGGAGTTTACAGCACT 1179
QY 1246 ATCAATGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
DB 1180 GTGAAAAGATTAATAATATCTGGAAGACATGATTTTATCTGTTGGTTGAA 1239
QY 1306 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1365
DB 1240 AAGCATGTGGAAGTCAAGCATTTTGGCTTATGCTCTTGTCTGATCTTTAACAAGCGA 1299
QY 1366 CCAGAGTGGGACAAAGCATTCAGGAGCTGAGCTAGTGAAGAGTGAAGAGAGAGAGAGAG 1425


```
Db 1300 CCTTGTGGGAGACCCCATTTTGTGTCTCCGTGAAGTCAATGATGCTGGTGGTGAAGATGAT 1359
QY 1426 GCCATCTACACAGCTGACACCGCTGCTGAGGTGACACAAAGCCCGAGCTTGTG 1485
Db 1360 CAGCTGATACATCACTGCTGCTACTGATGATGAC---AAACCAAGACTTGGTA 1416
QY 1486 ATCTGCTGCTGAGAGCGGAACTTGTGACATGGGGACCCCTATGTATGCTGGCTGAG 1545
Db 1417 GTACTCGATACAGAAAGAAACCCCTCAAGATGGTAACTTTACACAGTGGCAGTGAAG 1476
QY 1546 TCGGTCAAGCTGACACACACCGGAGAGCGCAGAGTACAGACCGGAGACCTTGC 1605
Db 1477 TCGGTCAATTTTGGCATCGGTGCTCCCGCTCCACAGTACATCAAGAGTAATCATATGT 1536
QY 1606 TCAGGCTTGTGCTG 1620
Db 1537 GCCGATTTCTCATC 1551

RESULT 8
US-10-416-314-89/c
; Sequence 89, Application US/10416314
; Publication No. US20040082508A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: GANDHI, Ameena G.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: CHAWLA, Nandinder K.
; APPLICANT: SANJANMALA, Madhusudan M.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: LU, Yan
; APPLICANT: GRETZEN, Kimberly J.
; APPLICANT: BUREFORD, Neil
; APPLICANT: DING, Li
; APPLICANT: HARALIA, April J.A.
; APPLICANT: TANG, Y. Tom
; APPLICANT: BANDMAN, Olga
; APPLICANT: WARREN, Bridget A.
; APPLICANT: HONGHELL, Cynthia D.
; APPLICANT: LU Dyrung, Aina M.
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: LEE, Sally
; APPLICANT: XU, Yuming
; APPLICANT: YANG, Junming
; APPLICANT: LAL, Preeti G.
; APPLICANT: TRAN, Bao
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: KAREHT, Stephanie K.
; TITLE OF INVENTION: SECRETED PROTEINS
; FILE REFERENCE: PI-0287 USN
; CURRENT APPLICATION NUMBER: US/10/416,314
; CURRENT FILING DATE: 2003-05-08
; PRIOR APPLICATION NUMBER: US 60/247,505
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: US 60/249,642
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/249,824
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,824
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: US 60/254,305
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: US 60/256,448
; PRIOR FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PERL Program
; SEQ ID NO 89
; LENGTH: 2460
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incycle ID No: 71947526CB1
US-10-416-314-89

Query Match 15.3%; Score 283.6; DB 17; Length 2460;
Best Local Similarity 98.6%; Pred. No. 1e-72;
Matches 286; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1365 GCCAGATGGGAGCAAGCACTACCGGAGCGTGAAGTGAAGAGAGTGAAGAGAGCA 1424
Db 2460 GCCAGATGGGAGCAAGCACTACCGGAGCGTGAAGTGAAGAGAGTGAAGAGAGCA 2401
QY 1425 CGCATCTACACAGCTACACAGCCCTGCTGAGGTGACACAAAGCCCGAGACTTGT 1484
Db 2400 CGCATCTACACAGCTACACAGCCCTGCTGAGGTGACACAAAGCCCGAGACTTGT 2341
QY 1485 GATCCTGCGCTGAGGCGGAGAGCCCTGTGACAAATGGGGACCCCTATGTCATCGGCTGAG 1544
Db 2340 GATCCTGCGCTGAGGCGGAGAGCCCTGTGACAAATGGGGACCCCTATGTCATCGGCTGAG 2281
QY 1545 GTGCGTCAAGCTGCGCCACACACCGAGAGACCGCAGAGTACAGACGCGGAGAGACCTTGT 1604
Db 2280 GTGCGTCAAGCTGCGCCACACACCGAGAGACCGCAGAGTACAGACGCGGAGAGACCTTGT 2221
QY 1605 CTCAGGCTTGTGCTGTGCGCGCGAGGCGGAGACCGAGTACCAAGTGTCT 1654
Db 2220 CTCAGGCTTGTGCTGTGCGCGCGAGGCGGAGACCGAGTACCAAGTGTCT 2171

RESULT 9
US-10-108-260A-1210
; Sequence 1210, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108, 260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1210
; LENGTH: 2242
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1210

Query Match 7.7%; Score 143.2; DB 16; Length 2242;
Best Local Similarity 85.5%; Pred. No. 2.2e-31;
Matches 183; Conservative 0; Mismatches 28; Indels 3; Gaps 2;

QY 1646 AGTGTCTGAGGTAGAGGTCTCCCTGACTGAGCTGTCTGCGCAAGTGAATTTCT 1705
Db 466 AGTGTCTGAGGTAGAGGTCTCCCTGACTGAGCTGTCTGCGCAAGTGAATTTCT 524
QY 1706 GGGGGCTGGAATCCAGGTCAAGGGTGGCAGAGCGACGGTTG--GAATGAAACTTGC 1763
Db 525 GGGGGCTGGAATCCAGGTGGAAGGGTCAAGAGTGAAGTGGGAACATGGAACCTTGC 584
QY 1764 TGGAGGACACCTGAGTACTCTTAAAGCAATCCCGTGGCCAAATCAACAGCGATTGG 1823
Db 585 TGGAGGACACCTGAGTACTCTTAAAGCAATCCCGTGGCCAAATCAATACCAATTGG 644
QY 1824 ATACCTCAAGACACCTGAACCTTATCATGAGC 1857
Db 645 ATACCTCAAGACACCTGAACCTTATCATGAGC 678

RESULT 10
US-09-867-701-2605/c
```



```
; Sequence 2605, Application US/09867701
; Patent No. US2002013237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2605
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-701-2605
```

```
Query Match 6.8%; Score 126.6; DB 9; Length 390;
Best Local Similarity 87.9%; Pred. No. 9.9e-27;
Matches 138; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```
QY 1644 CAAGTCTGCTGGGTTAGGTTCTCCCTGAGCTGTCTCCGCAAGTGGCTTCTATTC
DB 357 CTAGCTCTGATTGGTTAGAGTCTCCCGAGGAGCTGTCTCCGCAAGTGGCTTCTATTC
QY 1704 CTGGGGGCTGCAATCCAGTCAAGGTTCCAGAGCGAGGTTGAATGGAATACTAGC
DB 297 GTGAGGCTCGAATCCAGGTCAGGTCGAAGGTTGCTGAGAGCGAGGTTGAATGGAATACTAGC
QY 1764 TGAAGACACCTGAGTACTCTTAAAGCAATCCCGGTG 1800
DB 237 TGAAGACACCAAGTACTCTTAAAGCAATCCCGGTG 201
```

```
RESULT 11
US-09-814-353-21183
; Sequence 21183, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21183
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559,
; LOCATION: 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571,
; LOCATION: 572, 573
```

```
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21183
```

```
Query Match 6.8%; Score 126.2; DB 10; Length 573;
Best Local Similarity 91.2%; Pred. No. 1.5e-26;
Matches 134; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
QY 1654 TGGGTTAGGTTCTCCCTGAGTGTGTTCTGCAAGTGGCTTCTATTCCTGGGGGCTC 1713
DB 202 TGGGTTAGGTTCTCCCGAGAGAGCTGTCTCCGCAAGTGGCTTCTATTCCTGGAGCTC 261
QY 1714 GAATCCAGTCAAGGTTCCAGAGGAGCGGTTGGAATGGAATACTAGTGAAGACAC 1773
DB 262 GAATCCAGTCAAGGTTCTGAGAGCGAGGTTGGAATGGAATACTAGTGAAGACAC 321
QY 1774 CTGAGTACTCTTAAAGCAATCCCGGTG 1800
DB 322 CCAAGTACTCTTAAAGCAATCCCGGTG 348
```

RESULT 12

```
US-09-814-353-17043
```

```
; Sequence 17043, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
```

```
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
```

```
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17043
; LENGTH: 572
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-17043
```

```
Query Match 6.7%; Score 124.6; DB 10; Length 572;
Best Local Similarity 90.5%; Pred. No. 4.3e-26;
Matches 133; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

```
QY 1654 TGGGTTAGGTTCTCCCTGAGTGTGTTCTGCAAGTGGCTTCTATTCCTGGGGGCTC 1713
DB 202 TGGGTTAGGTTCTCCCGAGAGAGCTGTCTCCGCAAGTGGCTTCTATTCCTGGAGCTC 261
QY 1714 GAATCCAGTCAAGGTTCCAGAGGAGCGGTTGGAATGGAATACTAGTGAAGACAC 1773
DB 262 GAATCCAGTCAAGGTTCTGAGAGCGAGGTTGGAATGGAATACTAGTGAAGACAC 321
QY 1774 CTGAGTACTCTTAAAGCAATCCCGGTG 1800
DB 322 CCAAGCACTCTTAAAGCAATCCCGGTG 348
```

RESULT 13

```
US-09-918-995-21270
```

; Sequence 21270, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21270
; LENGTH: 543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(543)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-21270

Query Match 6.7%; Score 124; DB 10; Length 543;
Best Local Similarity 75.0%; Pred. No. 6,4e-26;
Matches 195; Conservative 0; Mismatches 60; Indels 5; Gaps 3;

QY 1601 TCTGCTACGGCTTCTGCTGCGGCGGAGGAGCCAGCTGACCAAGTCTGCTGGGTTA 1660
DB 34 TCGGACAGGAGAGCTGCGCCCGCTGCGAGGTATGCTGAGTTGCTCAGTGCCAGCTGGGTTA 93
QY 1661 GGGTCTCCCTGCTGCTGAGTGTGCTGCGAGTGTGCTTATTCCTGGGCGCTCGATCA 1720
DB 94 GGGTCTCCAGCAACGAGCTGTGCTCAGCAAGTGTGCTCAATCAT-GGGGCTCGAAGCTG 152
QY 1721 GGTCAAGGGTGTGCGAGGAGCGAGCTTGG--AATGAAAATCTAGCTGGAGACCTCTGA 1777
DB 153 GGTGAAGGGTGTGCGAGGAGCGAGCTTGGAGAACCTGGAACTAGCTGAGAGACCTCTGA 212
QY 1778 GTACTCTTAAGCAATCCCGGTGCGCAATCAACAGCGGATTTGGATTAACCTCAAGACA 1837
DB 213 GTACTCTT-AAAGCAATCCCGGTGCGCAATCAACCTGCAATTTGGATTAATCGAGGCA 271
QY 1838 CCTGAAACCTTATCATGAGC 1857
DB 272 CCTGAAACCTTATCATGAGC 291
RESULT 14
US-09-864-761-2374
; Sequence 2374, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Neomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2374
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008173.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
US-09-864-761-2374

Query Match 6.3%; Score 117.8; DB 9; Length 452;
Best Local Similarity 81.8%; Pred. No. 4e-24;
Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 1644 CAAGTGTGCTGGGTAGGTTCTCCCTGACGTGAGTGTCTGGCAAGTGTCTTATTC 1703
DB 70 CTAGTGCCACTGGGTAGGTTCTCCCTGACCGAGCTGGTCTGGCAAGTGTCTCA-TC 128
QY 1704 CTGGGGGTCAATTCAGGTCAAGGGTGTGAGAGCGACGCTTGAATGAAAATCTGC 1763
DB 129 GTGGGGGTCAATTCAGATGAAGTGTCTTGAAGCATGTGTAATGAAAATCTGC 188
QY 1764 TGAAGACACCTGAGTACTCTTAAAGCAATCCCGTGGCCAAATCAACAGCCGATTGG 1823
DB 189 TGAAGACACCGAGTACTCTTAAAGCAATCTCTGTGTGATGAAGAGGAGGAGCTAGA 248
QY 1824 A 1824
DB 249 A 249
RESULT 15
US-10-027-632-191598
; Sequence 191598, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

```
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: RastSeq for Windows Version 4.0
; SEQ ID NO 191598
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-191598
```

```
Query Match 6.2%; Score 115.2; DB 13; Length 594;
Best Local Similarity 83.1%; Pred. No. 2.6e-23;
Matches 143; Conservative 0; Mismatches 28; Indels 1; Gaps 1;
```

```
QY 1649 GCTGCTGGGTTAGGTTCCCTGACTGAGTGTCTTG- GCAAGTGGCTTCTATTCTCTGG 1707
    |||||||
Db 2 GCCACTGGGTTAGGTTCCCTGACCGAGCTGGCTTCGCGCAAGTGGCGTCCATTCGTGG 61

QY 1708 GGGCTGGAATCCAGTCAAGGCTCGCAGGAGGACCGGTTGGAATGGAATACTAGCTGA 1767
    |||||||
Db 62 GGGCTCAATCCAGTCAAGGCTTCGTAAGCGATCTTGACACGAAATCTAGCTGA 121

QY 1768 GGAACCTGAGTACTTTAAAGCAATCCCGTGCGCCAAATCAACAGCCGAT 1819
    |||||||
Db 122 GGACACCGGAGTACTTTAAAGCAATCCCGATGTGTAGTAAGAAAGGGAGYT 173
```

Search completed: August 21, 2004, 06:02:57
Job time : 895 secs

Blank Sheet

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2004, 20:19:58 ; Search time 5014 Seconds
(without alignments)

11059.846 Million cell updates/sec

Title: US-10-055-624B-1

Perfect score: 1857

Sequence: 1 gfggaattgcccttcaaat.....cctgaaaccttatcatgagc 1857

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_estl:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vri:*

28: gb_gssl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1526.6	82.2	1785	29	AY415998 Homo sapi
2	1230	66.2	2740	11	AK052332 Mus muscu
3	1230	66.2	2993	11	AK048670 Mus muscu
4	1230	66.2	3635	11	AK031588 Mus muscu

5	1211.8	65.3	2933	11	AK038546
6	1129.6	60.8	1781	29	AY416000
7	1081	58.2	1780	29	AY415999
8	858.8	46.2	914	13	EX451348
9	802.4	43.2	847	13	BQ223248
10	770.4	41.5	1008	13	EX332767
11	698	37.6	700	14	CA389831
12	636.4	34.3	751	9	AU141870
13	604	32.5	604	14	CD722495
14	596.6	32.1	892	13	BUI94842
15	592.6	31.9	911	12	BG774124
16	556	29.9	971	10	BE732224
17	479	25.8	640	14	CF532390
18	461.8	24.9	635	13	BY749381
19	460.4	24.8	462	13	EX492100
20	451.8	24.3	649	10	BB621160
21	437.6	23.6	702	14	CK300575
22	434.8	23.4	2012	11	AK033663
23	434.8	23.4	3050	11	AK004905
24	434.8	23.4	3056	11	AK034622
25	432.6	23.3	590	14	CB266287
26	429.8	23.1	700	14	CB518691
27	426.8	23.0	1026	10	BF780247
28	420.8	22.7	544	14	CB457406
29	419.8	22.6	487	10	BF603821
30	418.4	22.5	1208	14	CD505029
31	391.6	21.1	884	13	EX843189
32	388	20.9	556	12	EG710633
33	381	20.5	654	10	BB650432
34	378.8	20.4	578	10	BB632348
35	362.2	19.5	961	10	BF134551
36	354.4	19.1	604	10	BB659297
37	349.8	18.8	688	12	BJ001717
38	344	18.5	937	10	BF786458
39	341.8	18.4	635	14	CD218989
40	320.4	17.3	744	13	EX843754
41	316.8	17.1	350	12	BG956483
42	310	16.7	578	14	CD735167
43	301	16.2	575	14	CB152012
44	295.4	15.9	829	13	EX846471
45	295	15.9	827	13	EX844834

ALIGNMENTS

RESULT 1	AY415998	1785 bp	DNA	linear	GSS 17-DEC-2003
LOCUS	AY415998	Homo sapiens THEA gene, VIRTUAL TRANSCRIPT, partial sequence,			
DEFINITION	AY415998	Genomic survey sequence.			
ACCESSION	AY415998	GI:39771958			
VERSION	AY415998.1	GI:39771958			
KEYWORDS	GSS.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Inferred nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
REFERENCE	2 (bases 1 to 1785)				
AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Direct Submission				

JOURNAL

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.

COMMENT

FEATURES

source Location/Qualifiers

1..1785

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

<1..>1785

/gene="THEA"

/locus_tag="HCM5744"

gene

ORIGIN

Query Match 82.2%; Score 1526.6; DB 29; Length 1785;
Best Local Similarity 93.5%; Pred. No. 3.8e-299;
Matches 1529; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 19 ATGATCCAGAAATGTCGAAATCACCTGCGACGGGGCTTGGCCCTCTGTGTTCTCCAAACCGC 78
DB |||||
1 ATGATCCAGAAATGTCGAAATCACCTGCGACGGGGCTTGGCCCTCTGTGTTCTCCAAACCGC 60
QY 79 ACATCCCGGAAGTCAGCCTTACGTCGGGGAACGACAGTGCATGCCAGACGCGAGGGA 138
DB |||||
61 ACATCCCGGAAGTCAGCCTTACGTCGGGGAACGACAGTGCATGCCAGACGCGAGGGA 120
QY 139 TACCGGAACCCACGAGGTGAGATGAGCGAGCTGGTCTGCCCTGCCACACCAACCAA 198
DB |||||
121 TACCGGAACCCACGAGGTGAGATGAGCGAGCTGGTCTGCCCTGCCACACCAACCAA 180
QY 199 CGTGTGAGCTGAGGCTGGGGCAGCTGCTCAAGTGGATTTGACACCAACGGCTTGCCTGTCC 258
DB |||||
181 CGTGTGAGCTGAGGCTGGGGCAGCTGCTCAAGTGGATTTGACACCAACGGCTTGCCTGTCC 240
QY 259 GCGGAGAGCAGCTGGCTGCCCTGTGTGTCACAGCTTCCATGATGACATCTATTGTGAG 318
DB |||||
241 GCGGAGAGCAGCTGGCTGCCCTGTGTGTCACAGCTTCCATGATGACATCTATTGTGAG 300
QY 319 CACACATTAGTGTGGCAAGTGTGAATATCAAGGCCAAGGTGAACCGGCCCTTCAAC 378
DB |||||
301 CACACATTAGTGTGGCAAGTGTGAATATCAAGGCCAAGGTGAACCGGCCCTTCAAC 360
QY 379 TCCAGCATGAGGTGGGCATCCAGTGGCTCGAGGACCTGTGCTCTGAGAAGCAGTGG 438
DB |||||
361 TCCAGCATGAGNN 420
QY 439 AATGTGTGAAGGCTTGGCCACCTTGTGGGCCCGCGAGAGATCACCAAGTGAAGCTG 498
DB |||||
421 NNN 480
QY 499 AAGCAGATCACCGCGGACAGAGAGAGATGGAGCAGTGTGGCGCTGAGCGC 558
DB |||||
481 AAGCAGATCACCGCGGACAGAGAGAGATGGAGCAGTGTGGCGCTGAGCGC 540
QY 559 GCGGCATGCGCTTGTCTATGACAGACCATCAAGGACCTCTTGGCCAACTGCGCCATT 618
DB |||||
541 GCGGCATGCGCTTGTCTATGACAGACCATCAAGGACCTCTTGGCCAACTGCGCCATT 600
QY 619 CAGGCGCATGAGAGAGAGACATGTAGCCCATGTGGCCGCTGAGAGAACCCGTGTG 678
DB |||||
601 CAGGCGCATGAGAGAGAGACATGTAGCCCATGTGGCCGCTGAGAGAACCCGTGTG 660
QY 679 GAGAGTGGAGCTGGTCTGCTCCCGACGCAATCACAGGGCAACACCTTTCGGGCG 738
DB |||||
661 GAGAGTGGAGCTGGTCTGCTCCCGACGCAATCACAGGGCAACACCTTTCGGGCG 720
QY 739 CAGATCATGCGCTGATGAGATGTGCCACCATTTGCAGCCAGCGCGCTCTGCCGTGCC 798
DB |||||
721 CAGATCATGCGCTGATGAGATGTGCCACCATTTGCAGCCAGCGCGCTCTGCCGTGCC 780
QY 799 CACCCTAGCTCAAGGCCATTCAATTTCCATTTCCAGGCCCTCCAGGTTCGCGGAC 858
DB |||||
781 CACCCTAGCTCAAGGCCATTCAATTTCCATTTCCAGGCCCTCCAGGTTCGCGGAC 840

QY 859 CGTCTGCTCTCAAGCCATCGTGAACAATGCCTTCAAAACATAGCATGAGGTGGCGCTG 918
DB |||||
841 CGTCTGCTCTCAAGCCATCGTGAACAATGCCTTCAAAACATAGCATGAGGTGGCGCTG 900
QY 919 TCGTGGAGGCTTATCGCCAGGAGGTGAGACCCAGCGGCCACATCAACAGTGCCTTT 978
DB |||||
901 TCGTGGAGGCTTATCGCCAGGAGGTGAGACCCAGCGGCCACATCAACAGTGCCTTT 960
QY 979 ATGACCTTTGTGCTCTGAGACGAGATGACCGCCCAAGTGTGCTGCCCTTCGGCCC 1038
DB |||||
961 ATGACCTTTGTGCTCTGAGACGAGATGACCGCCCAAGTGTGCTGCCCTTCGGCCC 1020
QY 1039 CAGCCCGCGATGCTGAGCGCGGTACCGAGAGGCCAGTGCAGAAAGAGATCCGCTG 1098
DB |||||
1021 CAGCCCGCGATGCTGAGCGCGGTACCGAGAGGCCAGTGCAGAAAGAGATCCGCTG 1080
QY 1099 GACAGAAGTACATCGTCTTGAAGCAGACAGAGTGCCTCTCCCTCCCTGGAC 1158
DB |||||
1081 GACAGAAGTACATCGTCTTGAAGCAGACAGAGTGCCTCTCCCTCCCTGGAC 1140
QY 1159 CTTAGCAACAGGTGTACTGAGCTACAATAACGTCTCTCTTGAAGATGCTTGGCC 1218
DB |||||
1141 CTTAGCAACAGGTGTACTGAGCTACAATAACGTCTCTCTTGAAGATGCTTGGCC 1200
QY 1219 AAGGACAACTGGGTGCTCTCGGAGATCAGTCAAGTCCGCTGCTGAGGAT 1278
DB |||||
1201 AAGGACAACTGGGTGCTCTCGGAGATCAGTCAAGTCCGCTGCTGAGGAT 1260
QY 1279 GAACAAGTTCCTCTCCATGAGATGCTGATGATGATGATGATGATGATGATGATGAT 1338
DB |||||
1261 GAACAAGTTCCTCTCCATGAGATGCTGATGATGATGATGATGATGATGATGATGAT 1320
QY 1339 CTGCTGCTCTCGGACCTGCTGCTGAGGCGCAGAGTGGGCAAGCACTACCGAGAGTGGAG 1398
DB |||||
1321 CTGCTGCTCTCGGACCTGCTGCTGAGGCGCAGAGTGGGCAAGCACTACCGAGAGTGGAG 1380
QY 1399 CTAGTGACAGGTAGAGAGGACGAGCCATCTACCACTGCTGAGGCGGAGGAGGAGGAG 1458
DB |||||
1381 CTAGTGACAGGTAGAGAGGACGAGCCATCTACCACTGCTGAGGCGGAGGAGGAGGAG 1440
QY 1459 GGTACACAAAGCCCGAGGACTTCGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1518
DB |||||
1441 GGTACACAAAGCCCGAGGACTTCGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1519 GGGGACCCCTATGTCATCGCTGAGTGGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1578
DB |||||
1501 GGGGACCCCTATGTCATCGCTGAGTGGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
QY 1579 GAGTACAGCGGAGAGACCTCTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1638
DB |||||
1561 GAGTACAGCGGAGAGACCTCTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
QY 1639 CTGACCAAGTGTGCT 1654
DB |||||
1621 CTGACCAAGTGTGCT 1636

RESULT 2

AK052332

LOCUS

DEFINITION

AK052332 2740 bp mRNA linear HTC 20-SEP-2003
Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched
library, clone:D33029M16 product: brown fat inducible thioesterase
2 mRNA, full insert sequence.

ACCESSION

AK052332

VERSION

AK052332.1 GI:26342570

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus;

REFERENCE

1 Carninci, P. and Hayashizaki, Y.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

Location/Qualifiers
1..2993
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C230001C11"
/db_xref="MGI:2415236"
/db_xref="taxon:10090"
/clone="C230001C11"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
116..1900
/notes="unnamed protein product; brown fat inducible thioesterase 2 mRNA (GB|AF416923, evidence: BLASTN, 100%, match=2641)
putative"
/codon_start=1
/protein_id="BAC33416.1"
/db_xref="GI:26339490"
/translation="MIONVGNHLRRGFASMFNSTRKSIHPESGDPPTMARGEYVR NPTESQMSQVLVPCHTNHRGELSIGQLLKWIDTTACLSAERHAGCPVTFASMDIYED HTISVGQVNIKAKVNFARFNSMEVGIQVSEDLCKSEKQSVCKALATFVAHRELKSV KLQVPLTEERTHEGVAERMRLLVYADTILKOLLTHCVIQDDLDKDCSNVPAEK TRVESVELVLPHEHAGQNTFGGCIIMAMENVATIAASRLCHAHPTLKALEMVHFRPG SOVGRDLVLAIVNNAFKHSMGVCEAVRQEAETORRHINSAPFTFVFLDKDDQPO KLPIRPOGEGERRREARSKRIKLRDKVLSCKOARVALSVDPDPSNOVLYSYN VSSIKTLMAKDNVLSVEISEVRLYLEEDFLSFHLEMVNVDAAOVFQILLSLRBP EWDKHSRSELVQVEDDDAIHYVIFALSGRQKQDFVILASRRKPCDNDGPYVAL RSVTLPHRHTPEYORGETLCSGFLWREGDQMTKVSYNQATPGFLNYVTINVSGLS SEFYNTPKACESFLDNRNDLAPSLQLT"

ORIGIN

Query Match 66.2%; Score 1230; DB 11; Length 2993;
Best local Similarity 85.8%; Pred. No. 8.9e-239;
Matches 1404; Conservative 0; Mismatches 220; Indels 12; Gaps 3;
QY 18 AATGATCCAGATGTCGGAATACCTCGACGGGCTTGCCCTCTGTGTCTTCTCAACCG 77
DB 115 AATGATTCAGATGTGGCAACACTTGGAGAGGGCTTCGCCTATGTCTCTATCG 174
QY 78 CACATCCCGAAGTCAGCTTACGTGGC-----GGAAACGACAGTGCATGGCAGCGG 131
DB 175 CACATCCCGAAGTCAATCTCCCATCCGAGTCTGGAGACCTCTACCATGGCAGAGG 234
QY 132 CGAGGGATACCGGAACCCGACGAGGTGCAGATGACGAGTGGTGTGCTGCTGCCACAC 191
DB 235 TGAAGGATACCGGAACCCGACGAGGTGCAGATGACGAGTGGTGTGCTGCTGCCACAC 294
QY 192 CAACCAAGTGTGAGTGCAGTGCAGCTCGGACAGTCTCAAGTGGATGACACACGGCTTG 251
DB 295 CAACCAAGTGTGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 354
QY 252 CCTGTCCGCGAGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 311
DB 355 CCTATCAGCGAGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 414
QY 312 TTTTGAGCACACATAGTGTGGACAGTGGTGAATATCAAGCCCAAGGTGAACCGGGC 371
DB 415 CTTCCACCATACCATAGTGTGGCCCAAGTGGTGAATATCAAGCCCAAGGTGAACCGGGC 474
QY 372 CTTCAACTCCAGATGGAGTGGGATCCAGTGGCTCGGAGGACCTGTGCTCTTGAA 431
DB 475 CTTCAACTCCAGATGGAGTGGGATCCAGTGGCTCTGAGGATCTGTGCTCTGAGAA 534
QY 432 GCAGTGAATGTGTGCAAGGCTTGGCCACCTTCGTGGCCCGCGAGAGATCAACAAGT 491
DB 535 GCAGTGGAGTGTGCAAGGCTTGGCCACCTTCGTGGCCCGCGAGGCTCTCAAGT 594
QY 492 GAAGTGAAGCAGATCAAGCCCGGACAGAGAGAGAGATGGAGCAGTGTGGCGGC 551

DB 595 GAAGCTGAAGCAGGTATCCCATTTGACCGAGGAGAGAGACTGAACATGGGTGGCGGC 654
QY 552 TAGAGCCCGGCGATCGGCTTGTCTATGAGACACCATCAAGGACCTCTGCGGCACTG 611
DB 655 TGAGCGCGGCGGTATGCGATGCTGCTATGAGACACCATCAAGAGTCTCTAATACCCACTG 714
QY 612 CGCCATTCAAGGCGATCTGAGAGCAGAGACTGTAGCCGATGCTGCGGCTCAGAGAGAC 671
DB 715 TGTATCCAGGACGATTGGA---CAAGGACTGCAGCAATATGTTGCCAGCCGAGAGAC 771
QY 672 CCGTGTGAGAGTGTGGAGTGTGCTCTGCTCCCTCCCAAGCAATCAACAGGCGCAACCTT 731
DB 772 CCGAGTGGAGAGTGTGGAGTGTGCTGCTCTCCTCAGCCCAATCATCAGGCGCAATACCTT 831
QY 732 TGGGGCCAGATCATGGCTGTGATGAGATGTGGCCACCATTTGCAGCCAGCGGCTCTG 791
DB 832 CGGGGACAGATCATGGCTTGGATGGAGATGTGGCCACCATTTGCAGCCAGCGGCTCTG 891
QY 792 CCGTGCACCCCTACGCTGAAGGCCATTGAAATGTTCCACTTCCGAGGCGCGTCCCAAGT 851
DB 892 TCAGCGCCACCTTACGCTCAAGGCCATCGAGATGTTCCATTTCCGAGGCGCGTCTCAGT 951
QY 852 CGCGCACCGTCTGCTCAAAGCCATCGTGAACAATGCCCTTCAACATAGCATGAGGT 911
DB 952 GGGGACCGTCTGCTCAAGGCCATCGTGAATAAGCCCTTAAAGCACAGCATGGAGT 1011
QY 912 GGGCGTGTGGTGGAGCCCTATCCGAGAGGCTGAGACCCACCGGCGGCACATCAACAG 971
DB 1012 GGGTGTGTGTGGAGCGGTACCGCCAGGAAGCTGAGACCCAGCGCGGCACATCAACAG 1071
QY 972 TGCCTTTATCACCTTTGCTGCTTGGACGAGATGACACCCCGAGTTGCTGCCCTGGAT 1031
DB 1072 CGCTTCATGACCTTCGTGCTCTGGAACAAGATGACACCGCTCAGAGTGCCTCGAT 1131
QY 1032 TCGGCCCGCGCGGATGCTGAGCGCGGTACCGAGAGGCCAGTCCCAAGAAGAAGT 1091
DB 1132 TCGTCCCGACCTGGAGAGGTGACCGCGCATACCGAGAGGCCAGTCCAGGAAGAAGT 1191
QY 1092 CCGCTCGACAGGAAGTACATCGTGTCTGTAGCAGACAGAGTGCCTCTCGTCCGTC 1151
DB 1192 CCGCTCGACAGGAATACCTTGTGCTCTGTAAGCAGCAGAGTGGCCCTGTCTGTC 1251
QY 1152 CTGGACCTTAGCAACAGAGTGTACCTGAGTCAATAAAGTCTCTCTCTTGAAGATGCT 1211
DB 1252 CTGGACCTTAGCAACAGAGTGTACCTGAGTCTACAAAGTGTCTCTCTGAGAGCGT 1311
QY 1212 TGTGGCCAAAGCAACTCGGTGTCTCTCGAGATCAGTGCAGTCCGCTGTACCTCT 1271
DB 1312 CATGGCCAAAGCAACTCGGTGTCTCTCGAGATCAGCGAGTCCGCTGTACCTCT 1371
QY 1272 GGAGATGACAAGTTCCTCTCTTCCATGAGATGAGTGTGATGAGATGAGAGCCCA 1331
DB 1372 AGAAGAGGAC---TTCCTCTCTTCACTTGGAGATGTTGTAATGTTGATGCGGCCA 1428
QY 1332 GGCCTTCTCTGCTGCTCTCGGACCTCGCTCAGAGCGCAGTGGGACAGCACTACCGGAG 1391
DB 1429 GGTCTTTCAGTGTCTGTCAGACCTTCGCGAGGACCGAGTGGGACAGCACTTACCGGAG 1488
QY 1392 CGTGGAGTGTGTCAGCAGGTAGACGAGGACGAGGACGAGTGGGACAGCACTACCGGAG 1451
DB 1489 TGTGGAGCTGTGTCAGCAGTGGATGAGATGAGCCATCTACCGAGTCTACGCGCCG 1548
QY 1452 CCTCGAGGTCAACAAGCCCGAGGACTTGTGATCTGTGCTGAGGCGGAGGAGCTT 1511
DB 1549 CTTGAGCGGGAACCAAGCCCGAGGACTTGTGATCTGTGCTCTAGGCGGAGGAGCTT 1608
QY 1512 TGAACAATGGGACCCCTATGTCATCGCGCTGAGTTCGCTGAGTCCGCGCACACCGGAG 1571
DB 1609 TGAACAATGGGACCCCTATGTCATTTGCCCTGAGGTGCGTTCAGCTGCCACGACCATGA 1668
QY 1572 GACGCGAGGTACAGACCGCGAGAGACCTCTGCTCAGGCTTCTGCTCTCTGCGCGAGG 1631

Db 1669 GACACCGGAATCCACAGTGGGAGACTCTCTGTTCCAGGCTTCTGTCGTGGCGTGAGGG 1728

Qy 1632 GGACACGAGTCACCAAG 1647
|||||

Db 1729 GGACACGAGTACTAG 1744
|||||

RESULT 4
AK031588
LOCUS
DEFINITION
Mus musculus 3635 bp mRNA linear HTC 18-SEP-2003
enriched library, clone:6030456P18 product:brown fat inducible
thioesterase 2 mRNA, full insert sequence.
AK031588
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
MEDLINE
10349636
REFERENCE
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
11042159
REFERENCE
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
11076861
REFERENCE
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3635)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N.,
Okazaki, Y., Saio, R., Saichoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES

source

location/Qualifiers

1..3635

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM:DB:6030456P18"

/db_xref="MGI:2394836"

/db_xref="taxon:10090"

/clone="6030456P18"

/sex="male"

/tissue type="testis"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="13 days embryo"

33..1817

/note="unnamed protein product; brown fat inducible
thioesterase 2 mRNA (GB|AF416923, evidence: BLASTN, 100%,
match=2641)
putative"

/codon_start=1

/protein_id="BAC27464.1"

/db_xref="GI:2632741"

/translation="MIQVNGHLRRGFASMFNSRTSKSIHPESGDPPTWAEGSGYR
NPTVQMSQLVPLCHTHNRGLFSIGLLKWDITTAQLSAERHNGCPVTASMDIYFD
HTISGVGVNIIKAKVNAFNSMEVIGIVSLEDLSEQSWVCALATFVAREUSKV
KLQVILPTESEKTEHVAERKRLVIAUTIKDILLTHCVIQQDLKDCSNMVAEK
TRVSEPLTEHVAERKRLVIAUTIKDILLTHCVIQQDLKDCSNMVAEK
SQGDRLVLAIVNNAFKHMEVGVCEYRQEAETRRHNSAFMTFLWLDKDDQP
KLPIRVPQEGEGRYREASRKIRLDKLVYCKQAEVALSVWDFSNVYLSYN
BSSKTLMAKNVLSVEISVRVLYILEDFLSFLEVMVNDVAQVFLDLRRP
WMDKHSVELVQVDEDDALYHVISPALSGNTKPDQFVILASRRKPCDNGDPVYGL
RSVTLPTHETPEYQRCETLCSGFLWRREGDMTKVSYNQATPGFLNYVTNVSGLS
SEFYNTFRACSFLLDNRNDLAPSLQL"

CDS

ORIGIN

Query Match 66.2%; Score 1230; DB 11; Length 3635;
Best Local Similarity 85.8%; Pred. No. 9.9e-239;
Matches 1404; Conservative 0; Mismatches 220; Indels 12; Gaps 3;

Qy 18 AATGATCCAGATGTCGGAATACCTCGGACGGGGCTTGGCGCTCTGTGTCTTCAACCG 77
Db |||||

Qy 32 AATGATTCAGATGTCGGAACCACTTGGGAAGGGCTTGGCTCTATGTTCTCTAATCG 91
Db |||||

Qy 78 CACATCCCGGAGTACGCTTACGTGGG-----GGGAAACAGAGTGCATGGCAGCG 131
Db |||||

Qy 92 CACATCCCGGAGTACATCTCCATCCGAGTCTCGGAGACCCCTCTACCATGGCAGAGG 151
Db |||||

Qy 132 CGAGGGATACCGGAACCCCGAGAGTGCAGATGAGCCAGCTGTGCTGCCCTGCCACAC 191
Db |||||

Qy 152 TGAAGGATACCGGAACCCCGAGAGTGCAGATGAGCGCTGTGCTGTGCTGCCCTGCCAC 211
Db |||||

Qy 192 CAACCAACGTTGGTGTAGCTGAGCGTTCGGCGAGCTCTCAAGTGGATTGACACCGGCTG 251
Db |||||

Qy 212 CAACCAACGTTGGTGTAGCTGAGCGTTCGGCGAGCTCTCAAGTGGATTGACACCGGCTG 271
Db |||||

Qy 252 CCTGTCCCGGAGGAGGACCTGCTGCCCTGCTGTCACAGCTTCCATGATGATCATCTA 311
Db |||||

Qy 272 CCTATCAGCGGAGGAGGATGCTGGCTGCTCCGCTCACAGCTCTATGATGATCATCTA 331
Db |||||

Qy 312 TTTTGACACACCATTTAGTGTGGCAAGTGGTGAATATCAAGGCCAAGGTGAACCGGCG 371
Db |||||

Db 332 CTTGACCATACATAGTGTGGCGCAAGTGGTGAATATCAAGCCCAAGGTGACCGGC 391
 QY 372 CTTCAACTCAGCATGAGAGTGGGCATCCAGGTGGCTCGGAGGACCTGTGCTCTGAGAA 431
 Db 392 CTTCAACTCAGCATGAGAGTGGGCATCCAGGTGGTCTCTGAGGATCTGTGCTCTGAGAA 451
 QY 432 GCAGTGGAAATGTGCAAGGCTTGGCCACCTTCGTGGCCCGCGGAGAGATCAACCAAGT 491
 Db 452 GCAGTGGAGTGTCTGCAAGGCTTGGCCACCTTCGTGGCCCGCGGAGAGTCTCCAAGT 511
 QY 492 GAAAGCTGAAGCAGATCAAGCCCGGAGAGAGAGAGAGAGATGAGCACAAGTGTGGCGGC 551
 Db 512 GAAAGCTGAAGCAGATCAAGCCCGGAGAGAGAGAGAGAGATGAGCACAAGTGTGGCGGC 571
 QY 552 TGAGCGCGCGCATGCGCTTGTCTATGAGAGACCATCAAGGACCTCTGCGCAACTG 611
 Db 572 TGAGCGCGCGCATGCGCTTGTCTATGAGAGACCATCAAGGACCTCTGCGCAACTG 631
 QY 612 CGCATTCAGGCGCATCTGAGAGCAGAGACTGTAGCGCATGTGCGCGCTGAGAGAC 671
 Db 632 TGTATCATCAGGAGCATTTGGA---CAAGGACTGAGCAATATGTCGCGCGAGAGAC 688
 QY 672 CCGTGTGGAGAGTGTGAGTGTGCTGCTCCCGAGCCCAATCAGGAGGCAACACCTT 731
 Db 689 CCGAGTGGAGAGTGTGAGTGTGCTGCTCCCGAGCCCAATCAGGAGGCAACACCTT 748
 QY 732 TGGGGGCGAGATCATGCGCTTGTGAGAGAGTGTGCGCACCATGCGAGCGCGCTCTG 791
 Db 749 CGGGGAGACAGATCATGCGCTTGTGAGAGAGTGTGCGCACCATGCGAGCGCGCTCTG 808
 QY 792 CCGTGGCCACCTACCTGAGGCGCATGAAATGTTCCACTTCGAGGCGCGTCCAGGT 851
 Db 809 TCAGCGCCACCTACCTGAGGCGCATGAAATGTTCCACTTCGAGGCGCGTCCAGGT 868
 QY 852 CGGCGACCGTCTGTGCTCAAAGCATCTGTAACATGCGCTTCAAATGAGGCGCGTCTGAGT 911
 Db 869 GGGGACCGTCTGTGCTCAAAGCATCTGTAACATGCGCTTAAAGCACAGCATGAGGT 928
 QY 912 GGGGCTGTGAGGCGCTATGCCAGGAGGTGAGACCCACCGGCGCCACATCAACAG 971
 Db 929 GGGTGTGTGTGAGGCGTACCGCAGGAGCTGAGACCCGCGCGGACATCAACAG 988
 QY 972 TGCTTTATGACCTTGTGCTGTCGAGCAGATGACAGCCCGCTGCTGCTGCTGCTGAT 1031
 Db 989 CGCTTTCATGACCTTGTGCTGTCGAGCAAGATGACAGCTCAGAGCTGCTGCTGCTGAT 1048
 QY 1032 TCGGCCCGGCGGATGTCGAGCGCGGTACCGAGGCGGCTGCGAGAGAGAGAGAT 1091
 Db 1049 TCGTCCCGGCTGAGAGGCTGAAACCGCGATACCGAGAGCCAGTCCAGAGAGAGAT 1108
 QY 1092 CGGCTGAGAGAGTACATGCTGTCTGTAAGCAGAGAGTGGCGCTCTCGTCC 1151
 Db 1109 CGGCTGAGAGAGTACATGCTGTCTGTAAGCAGAGAGTGGCGCTCTGTGCTCC 1168
 QY 1152 CTGGGACCTTAGCAACAGGCTGACCTGAGCTCAATAGCTCTCTCTGAGAGTCT 1211
 Db 1169 CTGGGACCTTAGCAACAGGCTGACCTGAGCTCAATAGCTCTCTCTGAGAGTCT 1228
 QY 1212 TGTGGCAAGGACACTGGGTGCTCTCTCGAGATCAGTCAGTCTGCGCTGTCACCTCT 1271
 Db 1229 CATGGCAAGGACACTGGGTGCTCTCTCGAGATCAGTCAGTCTGCGCTGTCACCTCT 1288
 QY 1272 GGAGATGACAAAGTCTCTCTCTCCATCGAGATGCTGAGTGGATGAGAGAGAGAG 1331
 Db 1289 AGAAGAGGAC---TTCTCTCTCTCTCTCTGAGATGCTGTAATGAGTGGAGAGAG 1345
 QY 1332 GGCCTTCTGCTCTCTGAGTCTGAGTCTGAGGCGGAGTGGAGAGAGAGAGAGAGAG 1391
 Db 1346 GGTCTTTGAGTCTGTCAGACCTCTGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1405
 QY 1392 CGTGGAGTCTGTCAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1451

Db 1406 TGTGGAGTGGTGCAGCAAGTGGATGAGGATGAGCGCATCTACCATCTATCAGCCCGC 1465
 QY 1452 CTTGCGGAGTTCACAAAGCCCGCAGGACTTCGTGATCCTGCGCTCGAGGCGGAGCCTTG 1511
 Db 1466 CTTGAGCGGGAACACAAAGCCCGCAGGACTTTGTGATCCTGCGCTCTAGGCGGAGCCTTG 1525
 QY 1512 TGCAATGGGAGCCCTTATGTCTATCGCGTGTGAGTGGTGCAGCTGCGCCACACACCGAGA 1571
 Db 1526 TGCAATGGGAGCCCTTATGTCTATCGCGTGTGAGTGGTGCAGCTGCGCCACACCGAGA 1585
 QY 1572 GAGCCAGAGTACAGACGCGAGAGAGCCCTCTGCTCAGGCTTCTGCGCTGCGCGAGGG 1631
 Db 1586 GACACCGGATACCAACGTCGGGAGACTCTCTGTTTTCAGGCTTCTGCTGTCGTGGG 1645
 QY 1632 GGACCACTGACCAAG 1647
 Db 1646 GGACCACTGACCAAG 1661

RESULT 5 AK038546

LOCUS AK038546 2933 bp mRNA linear HTC 19-SEP-2003
 DEFINITION Mus musculus adult male hypothalamus cDNA, RIKEN full-length
 enriched library, clone:A23003M23 product:brown fat inducible
 thioesterase 2 mRNA, full insert sequence.

ACCESSION AK038546
 VERSION AK038546.1 GI:26332636

KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 2 99279253
 10349636

AUTHORS

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374
 11042159

JOURNAL

MEDLINE
 PUBMED
 11042159

REFERENCE

3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

Shibata,K., Itoh,M., Aizawa,K., Hayatsu,N., Sugahara,Y., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
 Sumi.N., Ishii.Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Tanaka,T., Matsura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913
 11076861

JOURNAL

MEDLINE
 PUBMED
 11076861

REFERENCE

4 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

JOURNAL

MEDLINE
 PUBMED
 11076861

REFERENCE

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

6 (bases 1 to 2933)

JOURNAL

MEDLINE
 PUBMED
 11076861

Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES

Location/Qualifiers

1..2933

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:A230033M23"

/db_xref="MGI:2402974"

/db_xref="taxon:10090"

/clone="A230033M23"

/sex="male"

/tissue type="hypothalamus"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev stage="adult"

99..1838

/note="unnamed protein product; brown fat inducible
thioesterase 2 mRNA (GB|AF16923, evidence: BLASTN, 100%,
match=2641)
putative"

/codon_start=1

/protein_id="BAC30036.1"

/db_xref="GI:26332637"

/translation="MFSNRTSRKSIHPESGDPPTMAGEGRNPTVQMSQVLPCH
TNHGEISIGQLKWDITTALESAPRHAGCPVTASMDIYFDHTISYQVNVIAKTV
NRAFNSMEVGIQVVSDELCEKQWCKALATFVAHRELKVKLQVILPTESEKTE
HGVAAERRRMLVVDITKLLTHCVIODLDDKDCNNVPAEKTRVESVELVLPHPAN
HOGNTFGQIWMWENVATIAASRLCHAHPITLKAEMPHFRPSQVGDRLVLAIVNN
AFKHSMEGVCEVAYROAETQRHINSAFMTFVLDKDDQPKLPIWRPQPGERR
YRESARKILRDKLYLSCQAEVALSVPMWPSNVYLSYINVSLLTKMAKONWVL
SVLESEVLYLLESDFTLSFHELMVNVDAQVQLLSLIRRPWDKHYRVELVQVQ
DEDDAIYHVSALSNKTPQDFVILASRRKPCNDGDPYIALRSVILPTHTETPEYQ
RGETLCSGFLWREGDQMTKVSYYNQATPGELNVVTTNVSLSEFYNTFACBSFL
DNRNDLAPSLQL"

ORIGIN

Query Match 65.3%; Score 1211.8; DB 11; Length 2933;
Best Local Similarity 85.6%; Pred. No. 4.3e-235;
Matches 1387; Conservative 0; Mismatches 22; Indels 12; Gaps 3;
QY 33 CGGAATACACCTGCGAGCGGCTGGCTCTGTCTTCCACCGCACATCCCGAAGTC 92
DB 68 CTGCACCTTCTAGCAGGCGCTTGGCTCTATGTCTTCTAATCGCATCCCGAAGTC 127
QY 93 AGCCTTACGTGCGGGA-----ACGACAGTGGCATGGCAGCGGAGGATACCGGAA 146
DB 128 AATCTCCATCGGAGTCTGGAGACCTCTTACCATGGCAGAGGTGAAGATACCGGAA 187
QY 147 CCCACGGAGGTGAGATGAGCCAGTGGTGTGCTGCTGCCACACCAACACGTGGTGA 206

DB 188 CCCACGGAGGTGAGATGAGCCAGCTGTACTGCTGCCACACCAACACCGTGGGA 247
QY 207 GCTGAGCGTGGGCGAGCTCTCAAGTGGATTGACACCGGCTTGCCTTCGCGCGAGAG 266
DB 248 GCTGAGCATTTGACAGTGTCTCAAGTGGATCGACACCAAGCTGCTTATCAGCGGAGAG 307
QY 267 GCACGCTGGCTGCCCTGTGTGTCACAGCTTCCATGGATGATCATCTATTTGAGCACA 326
DB 308 GATGCTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 367
QY 327 TAGTGTGGACAAAGTGGTGAATATCAAGGCCAAGTGAACCGGCTTCAATCCAGCAT 386
DB 368 TAGTGTGGCCTAAGTGGTGAATATCAAGGCCAAGTGAACCGGCTTCAATCCAGCAT 427
QY 387 GGAGTGGGCTATCCAGTGGCTCGAGGAGCTGTGCTCTGAGAGCACTGGAATGTGTG 446
DB 428 GGAGTGGGCTATCCAGTGGCTCTGAGGATCTGTGCTCTGAGAGCACTGGAATGTGTG 487
QY 447 CAAAGCTTGGCCTTCTGCTGGCGCGGAGAGATCAACAGGTGAAGTGAAGCAT 506
DB 488 CAAAGCTTGGCCTTCTGCTGGCGCGGAGAGATCAACAGGTGAAGTGAAGCAT 547
QY 507 CACGCGCGGACAGAGAGGAGAGATGAGCAGATGTGCGGCTGAGCGCGGCGCAT 566
DB 548 CATCCATTGACCGAGGAGAGAGATGAACTGAACTGCTGCTGCTGCTGCTGCTGCTG 607
QY 567 GCGCTTGTCTATGAGACACCATCAAGGACTCTCTGCGCACTGCGGCTGAGGCGCA 626
DB 608 GCGCTTGTCTATGAGACACCATCAAGGACTCTCTGCGCACTGCGGCTGAGGCGCA 667
QY 627 TCTGAGAGCAGAGACTGTAGCGGCTGCGGCTGAGAGACCTGCTGCTGCTGCTGCTG 686
DB 668 TTTGGA---CAGGACTGAGCAATATGCTGCGCAGCCGAGAGAGATCCGAGTGGAGTGT 724
QY 687 GAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 746
DB 725 GAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 784
QY 747 GCGCTGATGAGAAATGTGGCCACCATTTGAGCAGCGGCTGCTGCTGCTGCTGCTGCT 806
DB 785 GCGTGGATGAGAAATGTGGCCACCATTTGAGCAGCGGCTGCTGCTGCTGCTGCTGCT 844
QY 807 GCTGAAGGCCATTGAATGTTCCATCTCGAGGCGCGCTCCAGGTGCGGCGCGCTGCTG 866
DB 845 GCTCAAGGCCATCGAGATGTTCCATCTCGAGGCGCGCTCCAGGTGCGGCGCGCTGCTG 904
QY 867 GCTCAAGGCCATCGTGAACAATGCTTCAACATAGCATGAGGTGGGCTGCTGCTGCTG 926
DB 905 GCTCAAGGCCATCGTGAACAATGCTTCAACATAGCATGAGGTGGGCTGCTGCTGCTG 964
QY 927 GCGCTATCGCAGGAGGTGAGACCCAGCGGCGCACTCAACAGTGCCTTTATGACCTT 986
DB 965 GCGCTATCGCAGGAGGTGAGACCCAGCGGCGCACTCAACAGTGCCTTTATGACCTT 1024
QY 987 TGTGCTCTGAGCAGCATGACCCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1046
DB 1025 GGTGCTCTGACAAAGATGACCCAGCTTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1084
QY 1047 CGATGCTGAGCGGCTACCGAGAGGCGCAGTGCACAAAGAGATCCCGCTGGACGAA 1106
DB 1085 AGAGGGTGAACCGCGATACCCGAGAGGCGCAGTGCACAAAGAGATCCCGCTGGACGAA 1144
QY 1107 GTACATCGTGTCTGTAAGCAGACAGAGTGCCTCTCCCTGCTGCTGCTGCTGCTGCT 1166
DB 1145 ATACTTGTGCTGCTGTAAGCAGAGAGTGCCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1204
QY 1167 CCAGGTGTACCTGAGCTACAAAGTCTCTCTCTCTGTAAGATGCTTGTGGCCAGGACAA 1226
DB 1205 CCAGGTATACCTGAGCTACTACAGCTGCTCTCTCTGTAAGAGCTCTATGGCCAGGACAA 1264
QY 1227 CTGGGTGTGCTCTCGAGATCAGTCAAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1286

1265	CTGGGTGCTCTCGTGGAGATCAGCGAGGTCCGGCTGTATCATCTTAGAGAGGAC---	TT	1321
1287	CCTCTCCTCCACATGGAGATGGTGTGCATGTGGATSCAGCCAGAGCCCTTCCTGCTGCT		1346
1322	CCTCTCCTTTTCACTTGGAGATGGTGTGAATGTGGATGCCGCCAGGCTTTTCAGCTGCT		1381
1347	CTCGGACCTCGGTTCAGAGGCCAGTAGTGGGACACAGCACTACCGGAGCGTGGAGCTAGTGCA		1406
1382	GTCAAGACCTCGCGAGGAGACAGAGTGGGACAAGCATTTACCGGAGTGTGGAGCTGGTGCA		1441
1407	GCAGGTAGACGAGACGACGCCATCTACCACGTCACACAGCCCTGCCCTCGGAGGTCAAC		1466
1442	GCAAGTGATGAGATGACGCGCATCTACACAGTCTACGCCCGCCCTGAGCGGGAAACAC		1501
1467	AAAGCCCAGGACTTCCTGTATCTCTGGCTCGAGCGGAAGCCTTGTGACAAATGGGGACCC		1526
1502	CAAGCCCAGGACTTTTGTATCTCTGGCTCTTAGSCGAAGCCTTGTGACAAATGGGGACCC		1561
1527	CTATGTCATCGCGCTGAGGTTCGTCACGCTGCCACACACCGAGAGAGCCGACAGTACAG		1586
1562	CTATGTCATTTGCCCTGAGGTTCGCTACGCTGCCACACACCATAGACACACCGGAATACCA		1621
1587	ACGCGGAGAGACCCCTCTGCTCAGGCTTCTTGCTCTGGCGGAGGGGACACAGCTGACCAA		1646
1622	ACGTGGGGAGACTCTCTGTTTCAGGCTTCTGTCTGTGGCGTGGGGGACACAGTACTAA		1681
1647	G	1647	
1682	G	1682	

[illegible]

```

Db 1014 CGTCCCGGAGGCTGAGCGCGATACCGAGAGCCAGTGCACGAGGAAGATC 1073
QY 1093 GCGCTGGAGAGGATGATCTGCTGCTGAGGAGAGAGAGTGCCTCTCCGTCGCC 1152
Db 1074 GCGCTGGAGAGGAATACCTGTGCTGCTGAGGAGGAGAGTGCCTCTGCTGCTCC 1133
QY 1153 TGGGACCTTAGCAACAGGCTGACCTGAGCTACAACTGCTCTCTCTGAAAGATGCTT 1212
Db 1134 TGGGACCTTAGCAACAGGCTGACCTGAGCTACTACAGCTGCTCTCTGAGAGGCTC 1193
QY 1213 GTGGGCAAGGACAACTGGGTGCTGCTCGGAGATCAAGTCAAGTCCGCTGTACACTGT 1272
Db 1194 ATGGGCAAGGACAACTGGGTGCTGCTCGGAGATCAGGAGGCTCGGCTGTACATCTTA 1253
QY 1273 GAGGATGACAACTGCTCTCTCCCTCCACATGGAGATGGTGGTGCATGTGATCGAGCCAG 1332
Db 1254 GAAGAGGAC---TTCTCTCTCTTTCACCTTGGAGATGGTGGTGAATGTGGATCCGCCAG 1310
QY 1333 GCTTCTCTCTCTCTCTCGGACCTGCTGAGAGGCGGAGAGTGGGACAAAGCACTACCGGAGC 1392
Db 1311 GTCTTTTCAGCTGCTGTACAGCTGCTGCGAGGAGACCAAGTGGGACATTAACCGAGT 1370
QY 1393 GTGGAGCTAGTGCACAGGTAGAGAGGAGAGCGCATCTACAGTCAACGAGCCCTGCC 1452
Db 1371 GTGGAGCTGGTGCAGCAAGTGGATGAGGATGAGGATGAGGATCTACCACTATCAGCCCGCC 1430
QY 1453 CTCGAGAGTGCACAAAGCCCGAGGACTTCGATGATCTGGCTCGAGGCGGAGGCTTGT 1512
Db 1431 CTGAGCGGGAACCAAGCGCCAGGACTTTGTGATCCTGGCTTAGGCGGAGGCTTGT 1490
QY 1513 GACATGGGGAGCCCTATGTCTATCGGCTGAGTGGTCACTGCTGCCACACACCGGAGAG 1572
Db 1491 GACATGGGGAGCCCTATGTCTATGCTGCTGAGTGGTCACTGCTGCCACACCACTGAG 1550
QY 1573 AGGCAGAGTACAGAGCGGAGAGAGCCCTCTGCTCAGGCTTCTGCTCGGCGGAGGG 1632
Db 1551 ACACCGGAATACCAAGTGGGAGAGCTCTCTTTCAGGCTTCTGCTGCGGTGAGGG 1610
QY 1633 GACCAAGCTGACCAAG 1647
Db 1611 GACCAGATGACTAAG 1625

```

```

RESULT 7
LOCUS AY415999
DEFINITION Pan troglodytes THEA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY415999
VERSION AY415999.1 GI:39771959
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (bases 1 to 1780)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering

```

```

FEATURES
source Location/Qualifiers
1..1780
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1780
/gene="THEA"
/locus_tag="HCM5744"

ORIGIN
Query Match 58.2%; Score 1081; DB 29; Length 1780;
Best Local Similarity 66.4%; Pred. No. 1.2e-208;
Matches 1087; Conservative 0; Mismatches 549; Indels 0; Gaps 0;

QY 19 ATGATCCAGAAATGTCGGAATCACCTGCGACCGGGCTTGGCTCTGTGTCTTCCAAACCC 78
Db 1 ATGATCCAGAAATGTCGGAATCANNTGCGACGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60
QY 79 ACATCCCGGAAGTCAGCCTTACGTGCGGGGAACGACAGTGCATGGCAGACGCGGAGGA 138
Db 61 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
QY 139 TACCGGAACCCACGAGGTGCAGATGAGCCAGTGGTGTCTGCCCTGCCACCAACCAA 198
Db 121 TACCGGAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180
QY 199 CGTGTGAGCTGAGCGTGGGAGCTGCTCAAGTGATGACACACGCGCTTGCCTGTGCC 258
Db 181 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 240
QY 259 GCGGAGAGCGAGCTGGCTGCCCTGTGTACAGCTTCCATGGATGACATCTATTTTGAG 318
Db 241 NCGGAGAGCGAGCTGGCTGCCCTGTGTACAGCTTCCATGGATGACNNNNNNNNNNNN 300
QY 319 CACACCATTTAGTTGGACAAGTGTGAATATCAAGGGCCAAAGTGAAACCGGCGCTTCAAC 378
Db 301 NNNNNNNNNNTGTTGGACAAGTGTGAATATCAAGGCCAAAGTGAAACCGGCGCTTCAAC 360
QY 379 TCCAGATGAGGTGGGATCCAGGTGGCTCGGAGGACCTGTGCTCTGAGAAGCAGTGG 438
Db 361 TCCAGATGGAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420
QY 439 AATGTGTGCAAGGCTTGGCCACCTTCGTGGCCCCCGGAGAGATCACCAAGTGAAGCTG 498
Db 421 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480
QY 499 AAGCAGATCACGCGCGGACAGAGAGAGAGATGGAGCAGATGTGTGGCGCTGAGCGC 558
Db 481 AAGCAGATCACGCGCGGACAGAGAGAGAGATGGAGCAGATGTGTGGCGCTGAGCGC 540
QY 559 CGGCGCATGGCGCTTGTCTATGACAGACACCATCAAGGACCTCTCGGCCAATCGCGCAT 618
Db 541 CGGCGCATGGCGCTTGTCTATGCGGACCATCAAGGACCTCTCGGCCAATCGCGCAT 600
QY 619 CAGGCGCATTCGGAGAGCAGAGCTGTAGCGCATGTGGCGCTGAGAGAGACCGGTGG 678
Db 601 CANNNNNAITCGAGAGCAGAGACTGTAGCCCATGTGGTCCGCTGAGAGAGACCGGTGG 660
QY 679 GAGAGTGGAGCTGCTCTGCTCCCAACGCAATCAACAGGGGCAACCTTTGGGGGG 738
Db 661 GAGAGTGGAGCTGCTCTGCTCCCAACGCAATCAACAGGGGCAACCTTTGGGGGG 720
QY 739 CAGATCATGGCTGATGGAGATGTGGCCACCATTTGAGCAGCGCGCTCTGCCGTGCC 798
Db 721 CAGATCATGGCTGATGGAGATGTGGCCACCATTTGAGCAGCGAGNNNNCTCTGCCGTGCC 780
QY 799 CACCTACGCTCAAGGCTATTGAAATGTTCACTTCCGAGGCGCTTCCAGCTCGCGGAC 858
Db 781 CACCTACGNNNAANACANNNAATGTTTCACTTCCGAGGCGCTTCCAGCTCGCGGAC 840
QY 859 CGTGTGTGCTCAAGGCTATCGTGAACATGCTTCAACATAGCATGAGGTGGCGGTG 918

```


QY 1425 CGCCATACACAGTCACACAGCCCTGCGCTCGAGGTACACAAAGCCCGAGACTTCGT 1484
 Db 661 CGCCATACACAGTCACACAGCCCTGCGCTCGAGGTACACAAAGCCCGAGACTTCGT 720
 QY 1485 GATCCTGGCTGAGCGGAGCCCTTGACAAATGGGGACCCCTATGTATCGCGCTGAG 1544
 Db 721 GATCCTGGCTGAGCGGAGCCCTTGACAAATGGGGACCCCTATGTATCGCGCTGAG 780
 QY 1545 GTCCGTACAGCTGCCACACACCGAGAGAGCCGACAGGTACACAGCGGAGAGACCCCTCG 1604
 Db 781 GTCCGTACAGCTGCCACACACCGAGAGAGCCGACAGGTACACAGCGGAGAGACCCCTCG 840
 QY 1605 CTCAGCTTCTCCCTGTCGCGGAGGGGACACAGCTGACCAAGT 1648
 Db 841 CTCAGCTTCTCCCTGTCGCGGAGGGGACACAGCTGACCAAGT 884

RESULT 9
 BQ223248
 LOCUS
 DEFINITION AGENCOURT_7551001_NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6066060
 5' mRNA sequence.

ACCESSION BQ223248
 VERSION BQ223248.1 GI:20404648
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM13343 row: f column: 13
 High quality sequence stop: 647.
 Location/Qualifiers
 1. 847
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6066060"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC_72"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

FEATURES
 source

ORIGIN
 Query Match 43.2%; Score 802.4; DB 13; Length 847;
 Best Local Similarity 98.3%; Pred. No. 2.9e-152;
 Matches 820; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
 QY 162 GATGAGCAGCTGGTGGCTGCCCTGCCACCAACCAAGTGTGAGCTGAGCTCGGGCA 221
 Db 11 GATGAGCAGCTGGTGGCTGCCCTGCCACCAACCAAGTGTGAGCTGAGCTCGGGCA 70
 QY 222 GCTGCTCAAGTGGATTGACACACAGGCTTGCTGTCGCGAGAGGACGCTGGCTGCC 281
 Db 71 GCTGCTCAAGTGGATTGACACACAGGCTTGCTGTCGCGAGAGGACGCTGGCTGCC 130
 QY 282 CTGTGTCACAGCTCCATGGATGACATCTATTGAGCACACCATAGCTGTGACAGT 341

Db 131 CTGTGTACAGCTTCATGGATGACATCTATTGTTGAGCACACCATTAGTGTGGACAAGT 190
 QY 342 GGTGATATACAGGCCAAGGTGAAACCGGGCTTCACTCCAGATGGAGGTGGGATCCA 401
 Db 191 GGTGAATATCAAGGCCAAGGTGAAACCGGGCTTCACTCCAGATGGAGGTGGGATCCA 250
 QY 402 GGTGCGCTCGGAGGACCTGTGTCTGTGAGAACAGTGGAAATGTGTGCAAGGCTTGGCCAC 461
 Db 251 GGTGCGCTCGGAGGACCTGTGTCTGTGAGAACAGTGGAAATGTGTGCAAGGCTTGGCCAC 310
 QY 462 TTCTGTGCGCCCGCCGAGAGATCACCAAGGTGAAGTGAAGCAGATCACGCCCGGACAGA 521
 Db 311 TTCTGTGCGCCCGCCGAGAGATCACCAAGGTGAAGTGAAGCAGATCACGCCCGGACAGA 370
 QY 522 AGAGGAGAGATGGAGCACAGTGTGGCGCTGAGCGCGCGCGCGCGCATGCGCTTGTATGC 581
 Db 371 AGAGGAGAGATGGAGCACAGTGTGGCGCTGAGCGCGCGCGCGCATGCGCTTGTATGC 430
 QY 582 AGACACCATCAAGGACCTCTCTGGCCAACTGCGCCCATTCAGGCGCGATCTGGAGAGACAGA 641
 Db 431 AGACACCATCAAGGACCTCTCTGGCCAACTGCGCCCATTCAGGCGCGATCTGGAGAGACAGA 490
 QY 642 CTGTAGCGCGCATGTGTGGCGCTGAGAACACCGCTGTGAGAGATGTGGAGCTGGTCTGCC 701
 Db 491 CTGTAGCGCGCATGTGTGGCGCTGAGAACACCGCTGTGAGAGATGTGGAGCTGGTCTGCC 550
 QY 702 TCCCCACGCCCAATCAACAGGCAACACCTTTGGGGGCCAGATCATGGCTTGGATGGAGAA 761
 Db 551 TCCCCACGCCCAATCAACAGGCAACACCTTTGGGGGCCAGATCATGGCTTGGATGGAGAA 610
 QY 762 TGTGCGCCCAATTCAGCGAGCGCGCTCTGCGCTGCCACCCCTACGCTGAAGGCCATTGA 821
 Db 611 TGTGCGCCCAATTCAGCGAGCGCGCTCTGCGCTGCCACCCCTACGCTGAAGGCCATTGA 670
 QY 822 AATGTTCCACTTCCGAGCGCGCTCCAGGTGGCGACCGTCTGTGTGCTCAAAAGCCATCGT 881
 Db 671 AATGTTCCACTTCCGAGCGCGCTCCAGGTGGCGACCGTCTGTGTGCTCAAAAGCCATCGT 730
 QY 882 GAACAATGCTTCAACATAGCATGGAGTGGCGGTGTGCGTGGAGGCTATCGCCAGGA 941
 Db 731 GAACAATGCTTCAACATAGCATGGAGTGGCGGTGTGCGTGGAGGCTATCGCCAGGA 790
 QY 942 GGCTGAGACCCACCGCGCG-CCACATCAACAGTGCCTTTATGACCTTTGTGGTCC 994
 Db 791 GGCTGAGACCCACCGCGCGCCACATCAACAGTGCCTTTATGACCTTTGTGGGCC 844

RESULT 10
 BQ2232767
 LOCUS
 DEFINITION BX332767 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
 cDNA clone CS0DC015YP21 5-PRIME, mRNA sequence.

ACCESSION BX332767
 VERSION BX332767.1 GI:30312153
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: sequ@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 352.f For
 more information about this cluster, see
[http://www.genoscope.cns.fr/](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DC015CH1Q1P1&cluster=352.f)
 Contact :
 Feng Liang Email : fliang@lifetech.com URL :

```
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DC015CH11QPL1.
FEATURES
    source
        Location/Qualifiers
            1..1008
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS0DC015P21"
                /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                /notes="1st strand cDNA was primed with a NotI-oligo (dR)
                primer. Five prime end enriched, double-strand cDNA was
                digested with Not I and cloned into the Not I and EcoR V
                sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
    Query Match          41.5%; Score 770.4; DB 13; Length 1008;
    Best Local Similarity 98.0%; Pred. No. 1e-145;
    Matches 836; Conservative 3; Mismatches 7; Indels 7; Gaps 6;
QY 19 ATGATCCAGATGTCGGAATACCTCGACGGGCTTGGCTCTGTGTTCTTCCACCGC 78
DB 118 ATGATCCAGATGTCGGAATACCTCGACGGGCTTGGCTCTGTGTTCTTCCACCGC 177
QY 79 ACATCCCGGAAGTCAGCTTACGTGCGGGAAACGACAGTGCATGCGAGCGGAGGGA 138
DB 178 ACATCCCGGAAGTCAGCTTACGTGCGGGAAACGACAGTGCATGCGAGCGGAGGGA 237
QY 139 TACCGGAACCCACGGAGGTGCAGATGAGCCAGCTGTGTGCTGCCCTGCCACCAACCAA 198
DB 238 TACCGGAACCCACGGAGGTGCAGATGAGCCAGCTGTGTGCTGCCCTGCCACCAACCAA 297
QY 199 CGTGTGAGCTGAGCTGCGGACGCTGCTCAAGTGGATTGACACAGGCTTGCTGTCC 258
DB 298 CGTGTGAGCTGAGCTGCGGACGCTGCTCAAGTGGATTGACACAGGCTTGCTGTCC 357
QY 259 GCGGAGAGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318
DB 358 GCGGAGAGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 417
QY 319 CACACCAATTAGTGTGGAACAGTGGTGAATATCAAGGCCAAGGTGAACCGGCGCTTCAAC 378
DB 418 CACACCAATTAGTGTGGAACAGTGGTGAATATCAAGGCCAAGGTGAACCGGCGCTTCAAC 477
QY 379 TCCAGCATGAGGTGGGATCCAGTGGCTCGAGGACCTGTGCTGTGAGAACAGTGG 438
DB 478 TCCAGCATGAGGTGGGATCCAGTGGCTCGAGGACCTGTGCTGTGAGAACAGTGG 537
QY 439 AATGTGTGCAAGGCTTGGCCACCTTCTGTGCGCCGCGAGAGATCACCAAGTGAAGCTG 498
DB 538 AATGTGTGCAAGGCTTGGCCACCTTCTGTGCGCCGCGAGAGATCACCAAGTGAAGCTG 597
QY 499 AAGCAGATACCGCGCGGACAGAGAGAGAGAGATGAGACACAGTGTGGCGGCTGAGCGC 558
DB 598 AAGCAGATACCGCGCGGACAGAGAGAGAGAGATGAGACACAGTGTGGCGGCTTAAACGC 657
QY 559 CGGGCATGCGCTGTCTATGACACACCATCAAG-GACCTCTGCGCAACTCGCGCAT 617
DB 658 CGGGCATGCGCTGTCTATGACACACCATCAAGGACCTCTGCGCAACTCGCGCAT 717
QY 618 TCAGGGCGATCTGAGAGCAGAGACTGTAGCGCATGTGTGCGGCTGAGAGACCGCTGT 677
DB 718 TCAGGGCGATCTGAGAGCAGAGACTGTAGCGCATGTGTGCGGCTGAGAGACCGCTGT 777
QY 678 GGAGAGTGTGAGTGTGCTTGTCTTCCCGCAATCACAGGGGCAACACCTTTTGGGGG 737
DB 778 GGAGAGTGTGAGTGTGCTTGTCTTCCCGCAATCACAGGGGCAACACCTTTTGGGGG 837
QY 738 CCAGATCATGCGCTGGATGAGATGTGGCCACCATTTGACGACGAGCGGCTTCCCGTGC 797
DB 838 CCAGATCATGCGCTGGATGAGATGTGGCCACCATTTGACGACGAGCGGCTTCCCGTGC 896
QY 798 CCACCTTACGCTGAAGCCATTGAATGTTTCCACTTCCGAGGCCGCTCCCGAGTGGCGGA 857
```

```
DB 897 CCACCTACGCTGAGG-CATTGAAATGTT-CATTCCGAGGCC--GTCCAGGTGGCGCA 952
QY 858 CCGTCTGGTGTCTCAA 872
DB 953 -CGTCTGGTGTCTCAA 966

RESULT 11
CA389831 700 bp mRNA linear EST 06-NOV-2002
cs103b04.y1 Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unamplified); cs Homo sapiens cDNA clone cs103b04
5', mRNA sequence.
CA389831
CA389831.1 GI:24720367
EST.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
    1 (bases 1 to 700)
    Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,
    Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
    Expressed sequence tag analysis of human RPE/choroid for the
    NEIBank Project: Over 6000 non-redundant transcripts, novel genes
    and splice variants
    Mol. Vis. 8 (4), 205-220 (2002)
    22103460
    12107410
COMMENT
    Contact: Wistow G
    Section on Molecular Structure and Function
    National Eye Institute
    6/331, NIH, Bethesda, MD 20892-2740, USA
    Tel: 301 402 3452
    Fax: 301 496 0078
    Email: graeme@helix.nih.gov
    Plate: 103 row: b column: 04
    Seq primer: M13RPI reverse primer (ABI).
    Location/Qualifiers
        1..700
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="cs103b04"
            /tissue_type="RPE/choroid"
            /dev_stage="Adult"
            /lab_host="EMDH10B"
            /clone_lib="Human Retinal pigment epithelium/choroid cDNA
            (Un-normalized, unamplified); cs"
            /notes="Organ: Eye; Vector: pCMVSPORT6; Two different donor
            eyes (75-80 years old) yielded approximately 600 mg of
            dissected RPE/choroid tissue. This in turn yielded 340 ug
            of total RNA and 7 ug of mRNA. A directionally cloned cDNA
            library in the pCMVSPORT6 vector was constructed at Life
            Technologies (Rockville, MD; now part of Invitrogen Corp),
            essentially following the protocols of the SuperScript
            Plasmid System (Invitrogen Corp.
            <http://www.invitrogen.com/>). The library code
            designation was cs. For this library, cDNA inserts were
            cloned into the NotI/MluI sites of the vector. EST
            analysis was performed on the unamplified library at the
            NIH Intramural Sequencing Center (NISC)."
```

ORIGIN

```
Query Match          37.6%; Score 698; DB 14; Length 700;
Best Local Similarity 100.0%; Pred. No. 4.1e-131;
Matches 698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 711 CAATCACCAGGCAACACCTTTGGGGCCAGATCATGCTGGATGAGATGTGGCCAC 770
DB 3 CAATCACCAGGCAACACCTTTGGGGCCAGATCATGCTGGATGAGATGTGGCCAC 62
```

```

QY 771 CATTGACGACGCGCTCTGCGTGGCCACCTACGCTGAGGCAATTGAAATGTCCA 830
Db 63 CATTGACGACGCGCTCTGCGTGGCCACCTACGCTGAGGCAATTGAAATGTCCA 122
QY 831 CTTCCGAGCCCGTCCGAGGTGGCGACCGTGTGTGTCTCAAGCAATGCTGAA 890
Db 123 CTTCCGAGCCCGTCCGAGGTGGCGACCGTGTGTGTCTCAAGCAATGCTGAA 182
QY 891 CTTCAAAATAGCATGAGGTGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 950
Db 183 CTTCAAAATAGCATGAGGTGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 242
QY 951 CCACCGGCGCCACATCAACAGTGCCTTTATGACCTTTGTGTGTGTGTGTGTGT 1010
Db 243 CCACCGGCGCCACATCAACAGTGCCTTTATGACCTTTGTGTGTGTGTGTGTGT 302
QY 1011 GCGCCAGTTGTCCTGCTGATTTGCGCCCGCAGCCCGGCGATGTGTGTGTGTGT 1070
Db 303 GCGCCAGTTGTCCTGCTGATTTGCGCCCGCAGCCCGGCGATGTGTGTGTGTGT 362
QY 1071 GCGCCAGTTGTCCTGCTGATTTGCGCCCGCAGCCCGGCGATGTGTGTGTGTGT 1130
Db 363 GCGCCAGTTGTCCTGCTGATTTGCGCCCGCAGCCCGGCGATGTGTGTGTGTGT 422
QY 1131 AGAGGTGCCCCCTCTCCGTCCTGCGACCTAGCAACCAAGTGTGTGTGTGTGT 1190
Db 423 AGAGGTGCCCCCTCTCCGTCCTGCGACCTAGCAACCAAGTGTGTGTGTGTGT 482
QY 1191 CGTCTCTCTTTGAAGATGCTTGTGGCCAGGACCAACTGGGTGTGTGTGTGTGT 1250
Db 483 CGTCTCTCTTTGAAGATGCTTGTGGCCAGGACCAACTGGGTGTGTGTGTGTGT 542
QY 1251 TCAGTGTGCGCTGTACCTCTGAGGATGACAAAGTTCCTCTCTCTCCACATGAGATGT 1310
Db 543 TCAGTGTGCGCTGTACCTCTGAGGATGACAAAGTTCCTCTCTCTCCACATGAGATGT 602
QY 1311 GGTGATGTGGATGACGACCCAGCCCTTCTGTGTGTGTGTGTGTGTGTGTGTGT 1370
Db 603 GGTGATGTGGATGACGACCCAGCCCTTCTGTGTGTGTGTGTGTGTGTGTGTGT 662
QY 1371 GTGGACAGACACTACCGAGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1408
Db 663 GTGGACAGACACTACCGAGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 700

```

RESULT 12 AUI41870

LOCUS AUI41870 THYR01 Homo sapiens cDNA clone THYR01001374 5', mRNA
DEFINITION sequence.

ACCESSION AUI41870
VERSION AUI41870

KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 751)

Ota,T., NishiKawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.

HRI human cDNA project
Unpublished (2000)

CONTACT: Takao Isogai
Genomics Laboratory

Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 252-0812, Japan

Tel: 81-438-52-3975
Fax: 81-438-52-3986

Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix

Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and

FEATURES
source
1..751
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="THYR01001374"
/tissue type="thyroid gland"
/clone_lib="THYR01"
/note="Vector: pME18SFL3"

ORIGIN

Query Match 34.3%; Score 636.4; DB 9; Length 751;
Best Local Similarity 97.8%; Pred. No. 1.4e-118;
Matches 654; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

```

QY 19 ATGATCCGAGATGTGCGAAATCACCTGCGACCGGGCTTGGCCCTCTGTGTTCTCCAAACCGC 78
Db 83 ATGATCCGAGATGTGCGAAATCACCTGCGACCGGGCTTGGCCCTCTGTGTTCTCCAAACCGC 142
QY 79 ACATCCCGGAAGTTCAGCCTTACGTGCGGGGAACGACAGTGCATGCGACGCGGAGGA 138
Db 143 ACATCCCGGAAGTTCAGCCTTACGTGCGGGGAACGACAGTGCATGCGACGCGGAGGA 202
QY 139 TACCGGAACCCCAACGAGGTGCAGATGAGCCAGCTGGTGTCTCCCTGCCACCAACCA 198
Db 203 TACCGGAACCCCAACGAGGTGCAGATGAGCCAGCTGGTGTCTCCCTGCCACCAACCA 262
QY 199 CGTGTGAGCTGAGCGTGGGCGAGCTGCTCAAGTGGATGACACACGCGCTTGCCTGTCC 258
Db 263 CGTGTGAGCTGAGCGTGGGCGAGCTGCTCAAGTGGATGACACACGCGCTTGCCTGTCC 322
QY 259 GCGGAGAGGACGCTGGCTGCCCTGTGTACAGCTTCCATGCATGACATCTATTTTGAG 318
Db 323 GCGGAGAGGACGCTGGCTGCCCTGTGTACAGCTTCCATGCATGACATCTATTTTGAG 382
QY 319 CACACCATTTAGTTGGACAAAGTGTGTGAATATCAAGGCCAAGGTGAACCGGCGCTTCAAC 378
Db 383 CACACCATTTAGTTGGACAAAGTGTGTGAATATCAAGGCCAAGGTGAACCGGCGCTTCAAC 442
QY 379 TCCAGCATGAGGTGGGCATCCAGGTGGCTCGGAGGACCTGTGCTCTGAGAGAGTGG 438
Db 443 TCCAGCATGAGGTGGGCATCCAGGTGGCTCGGAGGACCTGTGCTCTGAGAGAGTGG 502
QY 439 AATGTGTGCAAGGCGCTTGGCCACCTTGTGGGCCCGCGAGAGATCACCAAGGTGAAGCTG 498
Db 503 AATGTGTGCAAGGCGCTTGGCCACCTTGTGGGCCCGCGAGAGATCACCAAGGTGAAGCTG 562
QY 499 AAGCAGATCACGCCCGGACAGAGAGAGATGGAGCAGAGTGTGGCGGCTGAGCGC 558
Db 563 AAGCAGATCACGCCCGGACAGAGAGAGATGGAGCAGAGTGTGGCGGCTGAGCGC 622
QY 559 CGGCGCATGCGCTTGTCTATGCAGACACCATCAAGGACCTCTGGCCAACTGGCCATT 618
Db 623 CGGCGCATGCGCTTGTCTATGCAGACACCATCAAGGACCTCTGGCCAACTGGCCATT 682
QY 619 CAGGCGCATCTGG-AGAGCAGAGACTGTAGCCCGCATGTGTGGCGGTGAGAAGACCCGCTGT 677
Db 683 CAGGCGCATCTGGAAAGAAAGCANAGACTGTACCCGATGTGTGGCGGCTTANAAAAACCCGCTGT 742
QY 678 GGAGAGTGT 686
Db 743 NGAAAAATGT 751

```

RESULT 13

CD722495

LOCUS

DEFINITION

o10d11.y1 Human lacrimal gland, unamplified: o1 Homo sapiens cDNA

clone o10d11 5', mRNA sequence.

CD722495

ACCESSION

CD722495.1 GI:32273343

VERSION

EST.

CD722495 604 bp mRNA linear EST 26-JUN-2003
o10d11.y1 Human lacrimal gland, unamplified: o1 Homo sapiens cDNA
clone o10d11 5', mRNA sequence.

CD722495
CD722495.1 GI:32273343
EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 604)
 AUTHORS Dickinson, D., Laurie, G. and Wistow, G.
 TITLE Expressed sequence tag analysis of human lacrimal gland
 JOURNAL Unpublished (2002)
 COMMENT Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 10 row: d column: 11
 Seq primer: M13RPI reverse primer (ABI).
 FEATURES
 source
 1..604
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="oj10d11"
 /tissue_type="lacrimal gland"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /clone_lib="Human lacrimal gland, unamplified: oj"
 /notes="Organ: Eye; Vector: pCMVSPORT6; RNA was extracted from 2 human lacrimal glands. A directionally cloned cDNA library in the pCMVSPORT6 vector(life technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual
 (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor
 (5'-pGACTAGTCTAGATCGGAGGCGGCC(T)15-3'). EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."
 ORIGIN
 Query Match 32.5%; Score 604; DB 14; Length 604;
 Best Local Similarity 100.0%; Pred. No. 4.6e-112;
 Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 418 CTGTGCTCTGAGAGCAGTGAATGTGTGAAGCCCTTGGCCACTTCGTGGCCCGCCGA 477
 DB 1 CTGTGCTCTGAGAGCAGTGAATGTGTGAAGCCCTTGGCCACTTCGTGGCCCGCCGA 60
 QY 478 GAGATCACCAGGTGAAGCTGAAGCAGATCACGCCGCGGACAGAGGAGAGATGGAG 537
 DB 61 GAGATCACCAGGTGAAGCTGAAGCAGATCACGCCGCGGACAGAGGAGAGATGGAG 120
 QY 538 CACAGTGTGGCGGTGAGCGCGCGCATGCGCTTCTATGATGAGACACCATCAAGGAC 597
 DB 121 CACAGTGTGGCGGTGAGCGCGCGCATGCGCTTCTATGATGAGACACCATCAAGGAC 180
 QY 598 CTCTGCGCAACTGCGCAATTCAGGGCGATCTGAGAGCAGAGACTGTAGCCGATGGT 657
 DB 181 CTCTGCGCAACTGCGCAATTCAGGGCGATCTGAGAGCAGAGACTGTAGCCGATGGT 240
 QY 658 CCGGCTGAGAGACCCGTGTGAGAGTGTGAGTGTGCTGCTCCCGCCAGCAATCAC 717
 DB 241 CCGGCTGAGAGACCCGTGTGAGAGTGTGAGTGTGCTGCTCCCGCCAGCAATCAC 300
 QY 718 CAGGGCAACACTTTGGGGGCCAGATCATGCGCTTGGATGGAGATGTGGCCACCATTTGCA 777
 DB 301 CAGGGCAACACTTTGGGGGCCAGATCATGCGCTTGGATGGAGATGTGGCCACCATTTGCA 360
 QY 778 GCCAGCCGGCTCTCGCGTGGCCACCCCTACCGCTGAAGGCCAATTTGAAATGTTCCACTTCCGA 837
 DB 361 GCCAGCCGGCTCTCGCGTGGCCACCCCTACCGCTGAAGGCCAATTTGAAATGTTCCACTTCCGA 420

QY 838 GGCCCTCCAGCTCGGCGACCGCTGTGGTCTCAAAGCCATCGTGAACATGCTTCAAA 897
 DB 421 GGCCCTCCAGCTCGGCGACCGCTGTGGTCTCAAAGCCATCGTGAACATGCTTCAAA 480
 QY 898 CATAGCATGAGAGTGGCGGTGTGCGTGGAGGCTTATGCGCAGAGGCTGAGACCCACCG 957
 DB 481 CATAGCATGAGAGTGGCGGTGTGCGTGGAGGCTTATGCGCAGAGGCTGAGACCCACCG 540
 QY 958 CGCCACATCAACAGTGCCTTTATGACTTTGTGCTCTGACGACGATGACCGACCCAG 1017
 DB 541 CGCCACATCAACAGTGCCTTTATGACTTTGTGCTCTGACGACGATGACCGACCCAG 600
 QY 1018 TTGC 1021
 DB 601 TTGC 604
 RESULT 14
 BU194842
 LOCUS BU194842
 DEFINITION AGENCOURT_7970201 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6164904
 5', mRNA sequence.
 ACCESSION BU194842
 VERSION BU194842.1 GI:22708826
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 892)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM13522 row: m column: 01
 High quality sequence stop: 554.
 FEATURES
 source
 1..892
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6164904"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_72"
 /notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."
 ORIGIN
 Query Match 32.1%; Score 596.6; DB 13; Length 892;
 Best Local Similarity 97.5%; Pred. No. 1.9e-110;
 Matches 627; Conservative 0; Mismatches 14; Indels 2; Gaps 2;
 QY 39 TCACCTCGCAGCGGGCTTGGCCCTCTGTGTTCTCCAACCGCACATCCCGAGTCAAGCTT 98
 DB 126 TCCCATCTTTGGGGCTTGGCCCTCTGTGTTCTCCAACACACATCCCGAGTCAAGCTT 185
 QY 99 ACGTGCGGGGAACGACAGTGCCTGCGAGCGGAGGATACCGAACCCCGAGGT 158
 DB 186 ACGTGGGGGAACGACAGTGCCTGCGAGCGGAGGATACCGAACCCCGAGGT 245
 QY 159 GCAGATGAGCGAGTGGTGTGCTGCTGCCACACCAACAGTGGTGGTGGTGGTGGTGG 218

Db 246 GCAGATGAGCCAGCTGGTGTCTGCTGCCCTGCCACACCAACCAAGCTGGTGTGAGCTGAGCGTGG 305

Qy 219 GCAGCTGCTCAAGTGATTTGACACACAGCTTTCCTCTCCGCGGAGAGCGACGCTGGCTG 278

Db 306 GCAGCTGCTCAAGTGATTTGACACACAGCTTTCCTCTCCGCGGAGAGCGACGCTGGCTG 365

Qy 279 CCCCTGTGTCAAGCTTCCATGGATGACATCTATTTTGGAGCACACCATTTAGTGTGGACA 338

Db 366 CCCCTGTGTCAAGCTTCCATGGATGACATCTATTTTGGAGCACACCATTTAGTGTGGACA 425

Qy 339 AGTGTGTGAATATCAAGGCGCAAGGTGAACCGGGCTTCACTCCAGCATGGAGGTGGGAT 398

Db 426 AGTGTGTGAATATCAAGGCGCAAGGTGAACCGGGCTTCACTCCAGCATGGAGGTGGGAT 485

Qy 399 CCAGTGGCTCGGAGGACCTGTCTCTGAGAAGCAGTGGATGTGCAAGGCCCTTGGC 458

Db 486 CCAGTGGCTCGGAGGACCTGTCTCTGAGAAGCAGTGGATGTGCAAGGCCCTTGGC 545

Qy 459 CACCTTCTGGCCCGCCGAGAGATCACCAAGGTGAAGCTGAAGCAGATCACCGCCGGAC 518

Db 546 CACCTTCTGGCCCGCCGAGAGATCACCAAGGTGAAGCTGAAGCAGATCACCGTTGGGAC 605

Qy 519 AGAGAGAGAGATGAGGACAGTGTGGCGGCTGAGCGCCGCGCATGCGCTTGTCTA 578

Db 606 AGAGAGAGAGATGAGGACAGTGTGGCGGCTGAGCGCCGCGCATGCGCTTGTCTA 665

Qy 579 TGCAGACACCATCAAGGACCTCTCGGCCAACTGGC-CCATTTCAGGGCGATCTGGAGAGCA 637

Db 666 TGCAGACACCATCAAGGACCTCTCGGCCAACTGGCCCACTGGCCCATTCAGGGCGATCTGGAGAGCA 725

Qy 638 GAGACTGTAGCGCATGTGTGCC-GGCTGAGAGACCCGCTGTGG 679

Db 726 GAGACTGTAGCGCATGTGTGCCGGGCTGAGAAGACCCCTGTGTG 768

RESULT 15
 BG774124 911 bp mRNA linear EST 15-MAY-2001
 LOCUS 602662096F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4810631 5',
 mRNA sequence.

ACCESSION BG774124.1 GI:14044423

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 911)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LUCM1666 row: p column: 24

High quality sequence stop: 851.

Location/Qualifiers

1..911

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4810631"

/tissue type="choriocarcinoma"

/lab host="DH10B (phage-resistant)"

/clone lib="NIH_MGC_21"

/note="Organ: placenta; Vector: pORF7; Site: 1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

	Query Match	31.9%;	Score 592.6;	DB 12;	Length 911;
	Best Local Similarity	99.3%;	Pred. No. 1.2e-109;		
	Matches 595;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
Qy	1056	CGCGCGGTACCGAGAGCGCAGTGCAGAAAGAAAGATCCGCTCGGACAGGAGTACATCGT	1115		
Db	2	CGCGCGGTACCGAGAGCGCAGTGCAGAAAGAAAGATCCGCTCGGACAGGAGTACATCGT	61		
Qy	1116	GTCTGTAAAGCAGACAGAGGTGCCCTCTCCGTCCCTCGGACCCCTAGCAACAGGTGTA	1175		
Db	62	GTCTGTAAAGCAGACAGAGGTGCCCTCTCCGTCCCTCGGACCCCTAGCAACAGGTGTA	121		
Qy	1176	CCTGAGCTACAAATACGTTCTCTTGAAGATGCTTGGCCAGGACCACTGGGTGCT	1235		
Db	122	CCTGAGCTACAAATACGTTCTCTTGAAGATGCTTGGCCAGGACCACTGGGTGCT	181		
Qy	1236	GTCTCTCGGAGATCAGTCAAGTCCGCTCTACACTCTGGAGGATGACAAAGTTCCTCTCCTT	1295		
Db	182	GTCTCTCGGAGATCAGTCAAGTCCGCTCTACACTCTGGAGGATGACAAAGTTCCTCTCCTT	241		
Qy	1296	CCACATGAGATGTTGGTGCATGTGGATGCGCCAGGACCCCTCTCTGCTCGGACCT	1355		
Db	242	CCACATGAGATGTTGGTGCATGTGGATGCGCCAGGACCCCTCTCTGCTCGGACCT	301		
Qy	1356	CGCTCAGAGCGCAGAGTGGGACAGCACTACCGGAGCGTGGAGCTAGTGCAGCAGTAGA	1415		
Db	302	CGCTCAGAGCGCAGAGTGGGACAGCACTACCGGAGCGTGGAGCTAGTGCAGCAGTAGA	361		
Qy	1416	CGAGGACGACGCCATCTACACAGTCCAGGACCCCTCGGAGGTACACAAAGCCCA	1475		
Db	362	CGAGGACGACGCCATCTACACAGTCCAGGACCCCTCGGAGGTACACAAAGCCCA	421		
Qy	1476	GGACTTCGTGATTCCTGGCTCGAGCGGAAAGCTTTGACAAATGGGGACCCCTATGTGAT	1535		
Db	422	GGACTTCGTGATTCCTGGCTCGAGCGGAAAGCTTTGACAAATGGGGACCCCTATGTGAT	481		
Qy	1536	CGCGCTGAGGTCGGTCAAGTCCGACACACCGAGAGACGCCAGAGTACAGACGCGGAGA	1595		
Db	482	CGCGCTGAGGTCGGTCAAGTCCGACACACCGAGAGACGCCAGAGTACAGACGCGGAGA	541		
Qy	1596	GACCTCTGTGTCAGGCTTCTGCTCTGGCGGAGGGGACCCAGCTGACCAAGGTATCTCT	1654		
Db	542	GACCTCTGTGTCAGGCTTCTGCTCTGGCGGAGGGGACCCAGCTGACCAAGGTATCTCT	600		

Search completed: August 21, 2004, 00:59:07
 Job time : 5026 secs

FEATURES
 source

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2004, 21:09:05 ; Search time 150 Seconds

(without alignments)
6870.292 Million cell updates/sec

Title: US-10-055-624B-1

Perfect score: 1857

Sequence: 1 ggggaatgcctttcaaat.....cctgaacattatcatgagc 1857

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*

5: /cgn2_6/ptodata/2/ina/6C COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.6	3.5	504	4	US-09-252-991A-15035
2	60.4	3.3	939	4	US-09-252-991A-1680
3	58.4	3.1	609	4	US-09-252-991A-2523
4	58.4	3.1	1518	4	US-09-252-991A-2333
5	56.6	3.0	480	4	US-09-252-991A-1630
6	56.2	3.0	1509	4	US-09-149-476-179
7	56.2	3.0	1530	4	US-09-149-476-306
8	53	2.9	1098	4	US-09-252-991A-14407
9	53	2.9	1566	4	US-09-252-991A-14264
10	53	2.9	1908	4	US-09-252-991A-14353
11	50.6	2.7	647	3	US-09-328-111-475
12	49.2	2.6	975	4	US-09-252-991A-5935
13	49.2	2.6	2196	4	US-09-252-991A-5953
14	49.2	2.6	3036	4	US-09-252-991A-5934
15	48	2.6	270	4	US-09-252-991A-14378
16	47.8	2.6	48328	4	US-09-536-002-27
17	46.8	2.5	2588	4	US-09-480-017-3
18	46	2.5	489	4	US-09-540-236-716
19	45.6	2.5	1190	6	5223414-1
20	45.4	2.4	1432	4	US-09-620-312D-420
21	45	2.4	1131	4	US-09-252-991A-6825
22	45	2.4	2463	4	US-09-252-991A-6775
23	44.2	2.4	1308	4	US-09-252-991A-914
24	44.2	2.4	1449	4	US-09-252-991A-1050
25	44	2.4	1182	4	US-09-252-991A-5793
26	44	2.4	1335	4	US-09-252-991A-5877
27	43.6	2.3	7218	1	US-08-232-463-14

Sequence 16, Appl
Sequence 7525, Ap
Sequence 7599, Ap
Sequence 7952, Ap
Sequence 1332, Ap
Sequence 12123, A
Sequence 12394, A
Sequence 1061, Ap
Sequence 3067, Ap
Sequence 2, Appl
Sequence 1, Appl
Sequence 12263, A
Sequence 8, Appl
Sequence 458, App
Sequence 6, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 7, Appl

US-09-818-780-16
US-09-252-991A-7525
US-09-252-991A-7599
US-09-252-991A-7952
US-09-023-655-1332
US-09-252-991A-12123
US-09-252-991A-12394
US-09-489-039A-3061
US-09-489-039A-3067
US-09-810-268-2
US-09-810-268-1
US-09-252-991A-12263
PCT-US91-02626-8
US-09-489-039A-458
US-08-312-870-6
US-08-307-444A-8
US-08-587-389-8
US-08-307-444A-7

ALIGNMENTS

RESULT 1

US-09-252-991A-15035
; Sequence 15035, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15035
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15035

Query Match 3.5%; Score 65.6; DB 4; Length 504;

Best Local Similarity 49.3%; Pred. No. 3.7e-06;

Matches 201; Conservative 0; Mismatches 204; Indels 3; Gaps 1;

QY 688 GAGTGGTCTGCTCCCAACGCAATCACCAGGGAACACCTTTGGGGCCAGATCATG 747
Db 37 GTGCTATGACCCCGGCAAGGAACTTCTCCGGCAACGTCACGCGCGGCGCTGCTC 96
QY 748 GCCTGATGAGAAATGGCCACCATTCAGCAGCAGCGGCTCTGCCGCCACCTTACG 807
Db 97 AAGCTCTCAGCAAGTGGCTTTGCTGCGCAAGGCTACGCGGACGCTAGTGGTG 156
QY 808 CTGAAGCCATGAAATGTTCCATTCCGAGGCGGCTCCAGGTCGCGGACCGTCTGGTG 867
Db 157 ACCCTGCGGTGATCAGGTGATCTTCGCGAGCGGATCATTCGCGGCACTGTTGAC 216
QY 868 CTCAAGCCATCTGACATGCTTCAACATAGATGAGGCTGGCGGTGGTGAG 927
Db 217 TTCTCGCTCGGTGAACTACACCGGCGGCTTCGATGAGGTCGGGTGAGAGGTGATG 276
QY 928 GCCTATCGCAGGAGGTGAGACCCCGGCGGCAATCAACAGTGCCTTTATGACCTTT 987
Db 277 ACCGAGACATCCAC---GAGCGCAGCGTGGCCATACCAAGTGTCTTTCACCATG 333
QY 988 TGCTCTCGACGACATGATACCGCCAGTTCGTCCTGATTCGCGCCAGCCGCGG 1047
Db 334 GTGGCATGACGACGACCGCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 393

QY 1048 GATGGTACGGCGGTACCGAGAGCGAGTGCACAGAAAGAGATCCGC 1095
 Db 394 ATCGAGAAACGCGCTACGCCAGCGCTTGGCCCGCGGAGCAGCGC 441

RESULT 2

US-09-252-991A-1680
 ; Sequence 1680, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 1680
 ; LENGTH: 939
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-1680

Query Match 3.3%; Score 60.4; DB 4; Length 939;
 Best Local Similarity 49.0%; Pred. No. 6.9e-05;
 Matches 190; Conservative 0; Mismatches 196; Indels 2; Gaps 1;

QY 689 AGCTGTCTCTGCTCCCAACGCAATCACAGGGCAACACCTTTTGGGGCCAGATCATGG 748
 Db 308 AGGGGTATTCGGCCCAACCAATCACCAACACCTGTTGGCGGAGCTGGCTGG 367
 QY 749 CTGATGAGAAATGTGGCCACATTGAGCCAGCGGCTCTGCGTGCCCACTACGC 808
 Db 368 CTGATGAGAGAGGTCTCTGTTTCATCACCGGACCGGTTTCTGCGCGCTGCGTGA 427
 QY 809 TGAAGGCCATTGAAATGTTCCACTTCCGAGCGCGTCCAGTCCGCGGACCGTCTGGNGC 868
 Db 428 CCGTTTCACCGATCGCATGCACTTCAACACCGCATCCGCGCGGGCTCGATCGTGA 487
 QY 869 TCAAGGCCATC--GTGAACATGCTTCAACATAGCATGAGGTGGCGGTGTGGTGA 926
 Db 488 ACTGTCGACGGGTGGTCAACGTCCGCAACACCATCTCAAGTGGAGTTCAGGTGA 547
 QY 927 GGCCTATCGCAGAGGTGAGACCCAGCGCGGTGCGGAAAGGCCATACACGGCTTTCAGCTTGT 607
 Db 548 CTGGAAAGCATGTACGCCGACGGTTCGCGAAGGCCATACACGGCTTTCAGCTTGT 607
 QY 987 TGTGCTCTGACGAGATGACCCAGCCCGTGTCTGCTGATTCGCGGATTCGCGCCAGCCCGG 1046
 Db 608 TGCCATCGACAGCAAGAAAGCGCGGTTCGCTGCTCGCGGGTTCGCGGTACGCGAGC 667
 QY 1047 CGATGGTACGCGGCTACCGAGAGGCC 1074
 Db 668 GGCTTGACAGGCGAAGAAAGGCAACC 695

RESULT 3

US-09-252-991A-2523
 ; Sequence 2523, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 2523
 ; LENGTH: 609
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-2523

Query Match 3.1%; Score 58.4; DB 4; Length 609;
 Best Local Similarity 47.1%; Pred. No. 0.00018;
 Matches 213; Conservative 0; Mismatches 236; Indels 3; Gaps 1;

QY 656 TGCCGGCTGAGAAGACCCGTGTGGAGAGTGTGAGCTGCTGCTCCCAAGCAATC 715
 Db 128 TGCAACCCGGAACCAACCACTGTGATGACCGTGTGATGATCGGACATGGCAATT 187
 QY 716 ACCAGGGCAACACCTTTGGGGCCAGATCATGGCCCTGGATGGAGAATGTGGCCACATTG 775
 Db 188 TCTCCGGCAATGTCCATGGCGGACCCCTCTGGAAGTATCTCGACGAAGTGCCTATGCAT 247
 QY 776 CAGCCAGCGGCTCTGCGCTGCCACCTACCTACGCTGAAGCCATTGAAATGTCCACTTCC 835
 Db 248 GCGCCAGTCGCTATGCCGGTCTGCTACGTGTGACCCCTGTGGTGTGATCAGGTGATCTTCC 307
 QY 836 GAGGCCCTCCAGGTCCGGGACCGTCTGGTGTCTCAAAGCCATCGTGAACAAATGCTTCA 895
 Db 308 GCGAGCCGGTGTGTCGGCAACTGTGTGACCTTCTCGCTTCCGTCACTACACGGGC 367
 QY 896 AACATAGCATGAGGTGGCGGTGTGCGTGGAGGCCCTATCGCCAGAGGCTGAGACCCACC 955
 Db 368 GGACTTCCATGGAAATCGCGTCAAGGTTCATCCCGAGAACATCCGCGAGCAGTCCGT-- 425
 QY 956 GGCGCCACATCAACAGTGCCTTTATGACCTTTGTGTCTCTGACGCAATGACGAGCCCC 1015
 Db 426 -GGGTATACCAACAGCTGTTTCTTCACTTGTGCGCTCGATGACGAGCGCAAGCGCG 484
 QY 1016 AGTTGCTCCCTGATTCGCGCCCGACCGCGCGATGGTGGCGGGTACCGAGAGGCCA 1075
 Db 485 TGGCCGTCCGCCCTGAGCTGGAGCTGGAAACCGCGGAGCAGAGCGGCTTGGCCAGGCC 544
 QY 1076 GTCCGAGAAAGAGATCCGCTCGACAGGAAG 1107
 Db 545 AACAAAGCGCGCAGCTGGCGCGGAGCTGGAG 576

RESULT 4

US-09-252-991A-2393/c
 ; Sequence 2393, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 2393
 ; LENGTH: 1518
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-2393

Query Match 3.1%; Score 58.4; DB 4; Length 1518;
 Best Local Similarity 47.1%; Pred. No. 0.00023;
 Matches 213; Conservative 0; Mismatches 236; Indels 3; Gaps 1;

QY 656 TGCCGGCTGAGAAGACCCGTGTGGAGAGTGTGAGCTGCTGCTCCCAAGCAATC 715

Db 650 TGAACCCGGAACACCACTGCTGATGACCGTGTGATGATCGCGACATGGCCATT 591
QY 716 ACAGGGCAACACTTTGGGGGCGAGATCATGSCCTGGATGGAGAATGTGGCCACCATTTG 775
Db 590 TCTCCGGCAATGCTCATGGCGGACCCCTCTGAAGTATCTCGACGAAAGTGGCTATGCA 531
QY 776 CAGCGAGCGGCTCTGCGGTGCCACCTACGCTGAAGGCATTGAAATGTTCCATTCC 835
Db 530 GCGCCAGTCGCTATGCGCGTCTGCTGATGACCTGTGCTGGATCAGGTGATCTTCC 471
QY 836 GAGCGCCGTCCTCCAGGTGCGGCGACCTGCTGCTCAAGACCATCTGTAACAATGCTTCA 895
Db 470 GCGAGCGGTGCTGCTGCGGCACTGCTGACCTTCTCGTTCGTTCACTACCGGGC 411
QY 896 AACATGATGAGGTGGGTGCTGCTGAGAGGCTATCCCGAGGAGGCTGAGACCCACC 955
Db 410 GGAATTCCTGGAATCGGCGTCAAGTCTATCCGAGAACATCCGCGAGCAGTCCGT-- 353
QY 956 GCGCCACATCAACAGTCCCTTTATGACCTTTGCTGCTGAGCGGATGACGAGCCGCC 1015
Db 352 -GCGTCATACCAACAGTGTCTTCCACCAATGCTGCGCTTGGATGAGCGGCAAGCCGG 294
QY 1016 AGTTGCTGCTGCTGATTCGCGCCCGACCGCGGATGCTGAGCGCGGTACCGAGAGGCCA 1075
Db 293 TGCGCGTCCGCGCTGAGCTGGAACCGCGAGAGCGCGCTTCGCCAGGCC 234
QY 1076 GTCCCAAGAAAGATCCCGCTCGACAGGAAG 1107
Db 233 AACAAAGCGCCAGCTGCGCGGAGGTGGAG 202

RESULT 5

US-09-252-991A-1630
; Sequence 1630, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1630
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1630

Query Match 3.0%; Score 56.6; DB 4; Length 480;
Best Local Similarity 49.6%; Pred. No. 0.00043;
Matches 173; Conservative 0; Mismatches 174; Indels 2; Gaps 1;

QY 689 AGTGGTCTGCTCCCGACGCCAATCACCAGGGGCAACACTTTGGGGGCGCAGATCATGG 748
Db 114 AGCGGTATTCCCGCCACCACCAATCACCACACCTCTTTCGGCGGAGCTGCGGTGG 173
QY 749 CTTGGATGGAGATGTGGCCACCATTCGACGAGCGGCTCTGCGGTGCCACCCCTACGC 808
Db 174 CTTGGATGGAGAGTCTGCTTATCACCAGCGACCCGCTTCTGCGCGCTGCGGTGTA 233
QY 809 TGAAGCCATTGAATGTTCCATCTCCGAGCGGCTCCAGGTCGGGACCGCTGCTGTC 868
Db 234 CCGTTTCCACCGATCGCATCTTCAACACCGATCCCGGGGGGCTCGATCGTCGA 293
QY 869 TCAAGGCATC--GTGAACCAATGCTTCAACACATAGCATGAGGTGGCGGTGTGCGTGA 926
Db 294 ACTGGTCGAGCGGTGTGTCACAGTCGCGCAACACCATCTCTCAAGGTGAGGTGAGGTGA 353

QY 927 GGCTATCCGAGGAGGCTGAGACCCACCGGCGCCACATCAACAGTGCCTTTATGACCTT 986
Db 354 CTGGAAGACATGTACGCGGACGCTCGGAGAAAGGCCATACACGGGCTGTTCACTTCGT 413
QY 987 TGTGGTCTTGGAGCGCAGATGACCGGCCCGGAGTTGCTGCTGCTGCTGCTGCTGCTGCTG 1035
Db 414 TGCCATCGACGAGAAAGCGGCGGTTCCGGTCTCTGCGCGGGTTCCTGCGCGGGTTCCTG 462

RESULT 6

US-09-149-476-179
; Sequence 179, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596

Db 654 TGATCCACCTGGTGGGCGCTTACAGCTGCACCCCTGCACGGCTTTGTGCACGAGGTGTGA 713
QY 224 TGCTCAAGTGATGACACACGCTTGCCTGTCCGGGAGAGCAGCTGGCTGGCCCT 283
Db 714 CATTGAGCTCATGATGAGTCCCGGATCGTGGCTGCACGCACTGCAAGACCAACA 773
QY 284 GTGTACAGCTTCATGATGACATCATTTTGTGACACACCATTTAGTGTGGCAAGTGG 343
Db 774 TCGTCACAGCTTCGTTGAGCGCCATTATTTTATGACACAGATCAGAAAAGGTGGTCA 833
QY 344 TGAATATCAAGGCAAGGTGAACGGGCGCTTCACTCCAGATGAGAGTGGGATCCAGG 403
Db 834 TCACCATCTCGGACGATGACCTTACAGAGCAATAAGTCCATGGAGATCGAGGTGTGG 893
QY 404 TGGCCTCGGAGGACCTGT-----GCTCTCAGAGCAGTGAATGTGTGCAAGGCTTGG 457
Db 894 TGGACGCGGACCTGTGTGTGAGACGCTCTCAGAAAGCGCTACCGGGCGCGCAGTGCCTTCT 953
QY 458 CCACCTTCTGTGGCGCGCGGAGATCACCAGGTGAAGCTGAAGCAGATCACGCGCGGGA 517
Db 954 TCACCTACGTGTGCTGAGCGAGGAGGAGTGCCTGTGCTGTGCCCCAGCTGGTGGCCG 1013
QY 518 CAGAGAGGAGATGAGGACACAGTGTGGCGCTGAGCGCGCGCGCATGC 568
Db 1014 AGACCGAGGACGAGAGAGCGCTTTGAGGAAGGCAAGGCGGTACCTGC 1064

RESULT 7

US-09-149-476-306
; Sequence 306, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: F2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888

	; EARLIER APPLICATION NUMBER: 60/057,650	
	; EARLIER FILING DATE: 1997-09-05	
	; EARLIER APPLICATION NUMBER: 60/056,884	
	; EARLIER FILING DATE: 1997-08-22	
	; EARLIER APPLICATION NUMBER: 60/057,669	
	; EARLIER FILING DATE: 1997-09-05	
	; EARLIER APPLICATION NUMBER: 60/049,610	
	; EARLIER FILING DATE: 1997-06-13	
	; EARLIER APPLICATION NUMBER: 60/061,060	
	; EARLIER FILING DATE: 1997-10-02	
	Query Match 3.0%; Score 56.2; DB 4; Length 1530;	
	Best Local Similarity 47.9%; Pred.No.0.00073;	
	Matches 197; Conservative 0; Mismatches 208; Indels 6; Gaps 1;	
Qy	164 TGAGCCAGCTGTGTCCTGCCACACCACCAACCGTGTGAGCTGAGCGTCGGGCAGC 223	
Db	745 TGAATCACCTGTGTGGGGCCTTCAGACTGCACCCCTTTGTGSCAGGAGGTGTGA 804	
Qy	224 TGCTCAAGTGATTGACACCACGGCTTGCTGCCGGGAGAGCACGCTGGCTGCCCT 283	
Db	805 CCAATGAAGCTCATGGATGAGTGC CGGGATCGTGGCTGCACGCCACTGCAAGNCAACA 864	
Qy	284 GTGTCAAGCTTCCATGTGATGCAATCATTTTTGAGCACACCATAGTGTGGACAAGTGG 343	
Db	865 TCGTCAAGCTTCCGTGGAGCCATAAATTTTCATGACAGATCAGAAAAGGTGCGCTCA 924	
Qy	344 TGAATATCAAGGCCAAGGTGAACCGGGCCTTCAACTCCAGCATGGAGTGGCATCCAGG 403	
Db	925 TCACCATCTCGGAGCGCATGACCTTCACGAGCAATTAAGTCCATGGAGATCGAGTGTGG 984	
Qy	404 TGGCCTCGGAGGACCTGT-----GCTCTCAGAAGCAGTGAATGTGTGCAAGGCTTTGG 457	
Db	985 TGGACGCGGACCCTGTTGTGGACAGCTCTCAGAAGCGCTACCGGGCGCCAGTGCCTTCT 1044	
Qy	458 CAACCTTCGTGGCGCCGCGAGAGATCACCAAGGTGAAGCTGAAGCAGATCACGGCGGGA 517	
Db	1045 TCACCTACGTGTGCGCTGAGCCAGGAAAGCGAGTGTGCTGTGCCCGGCTGTGCTGCGCG 1104	
Qy	518 CAGAAGAGGAGAAGATGAGACACAGTGTGGCGCTGAGCGCGCGCGCATGC 568	
Db	1105 AGACCGAGGACGAGAGAAGCGCTTTGAGAAAGCAAGGCGGTACTCTGC 1155	
RESULT 8		
US-09-252-991A-14407	; Sequence 14407, Application US/09252991A	
	; Patent No. 6551795	
	; GENERAL INFORMATION:	
	; APPLICANT: Marc J. Rubenfield et al.	
	; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD	
	; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS	
	; FILE REFERENCE: 107196.136	
	; CURRENT APPLICATION NUMBER: US/09/252,991A	
	; CURRENT FILING DATE: 1999-02-18	
	; PRIOR APPLICATION NUMBER: US 60/074,788	
	; PRIOR FILING DATE: 1998-02-18	
	; PRIOR APPLICATION NUMBER: US 60/094,190	
	; PRIOR FILING DATE: 1998-07-27	
	; NUMBER OF SEQ ID NOS: 33142	
	; SEQ ID NO 14407	
	; LENGTH: 1098	
	; TYPE: DNA	
	; ORGANISM: Pseudomonas aeruginosa	
US-09-252-991A-14407		
Query Match	2.9%; Score 53; DB 4; Length 1098;	
Best Local Similarity	49.1%; Pred.No.0.0036;	
Matches 140; Conservative	0; Mismatches 145; Indels 0; Gaps 0;	
Qy	321 CACCATTAGTGTGGACAAGTGTGTAATATCAAGGCCAACCGGCGCTTCAACTC 380	
Db	445 CAACATGGGCTTCGCCGAAGCGGTGGACTGTCCGGTGTATCTGTGTCGCGATATCGACCG 504	

Qy	381	CAGCATGGAGGTGGGCATCCAGGTGGCTTCGGAGGACCTGTGCTCTGGAAGCAGTGGAA	440
Db	505	CGGGGGGTCTTCGCCCACTGGTCGGCACCTCTGGAACCTGCTCTCGACTCGAGCGGGA	564
Qy	441	TGTGTGCAAGGCTTGGCCACCTTCGTGGCCCGCGAGAGATCAACAAGTGAAGCTGAA	500
Db	565	CGGGGTCAAGGGCTTCGTATCAATTCGTTCGCGCGCACATCGCCCTGCTGCAACCCGG	624
Qy	501	GCAGATCACGGCGCGGACAGAGAGAGATGGAGCACAGTGTGGCGCTGAGCGCG	560
Db	625	CCTCGACTGGCTTGAGGACACGACCGCGAAGCCGTCTCGCGTGTGCCCTATGTCTAG	684
Qy	561	CGCATGCGCCTTGCTCTATGAGACACCATCAAGGACCTCTCGGC	605
Db	685	CGACCTGCACCTTGAGGGCGGAGGAGCCCATGCACACGCGCCAGGC	729

```

RESULT 9
US-09-252-991A-14264/c
; Sequence 14264, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14264
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14264

```

Query Match	2.9%;	Score 53;	DB 4;	Length 1566;
Best Local Similarity	49.1%;	Pred. No. 0.004;		
Matches 140;	Conservative 0;	Mismatches 145;	Indels 0;	Gaps 0;
QY 321	CACCATAGTGTGGACAAGTGTGTGATATCAAGGCCAAGGTGACCGGCGCTTCAACTC	380		
Db				
QY 1099	CAACATGGGCTTCGCGAAGCGGTGGACTGTCGGTGATCTGCTGCCGATATCGACC	1040		
Db				
QY 381	CACCATGGAGGTGGGCATCCAGGTGGGCTCGGAGGACCTGCTCTGAGAACGATGGAA	440		
Db				
QY 1039	CGCGGGGCTTCGCGCCACTGTGTCGGCACCTTGGAACTCTCTCGACATCGGAGCGGA	980		
Db				
QY 441	TGTGTGCAAGGCGCTTGCGCCACTTCTGTGGCGCCGCGAGAGATCACCAAGGTGAAGTGAA	500		
Db				
QY 979	GCGGGTCAAGGGCTTCGTCTCAATCGCTTCGCGGCGCATCGCCCTGCTGCAACCCGG	920		
Db				
QY 501	GCAGATACGCCCGGACAGAGAGGAGAGATGGAGCACAGTGTGCGCGCTGAGGCGCG	560		
Db				
QY 919	CCTCGCATGGCTGGAGGACGCGACCGCAAGCGGTGCTCGGCGTGTGCGCTATGTGAG	860		
Db				
QY 561	GCGCATGCGCCTTGTCTATGCAGACACCATCAAGGACCTCTCGC	605		
Db				
QY 859	CGACCTGCACCTGGAGCGGAGGACGCGCATTCGACAGCGCGCAGGC	815		
Db				

```

RESULT 10
US-09-252-991A-14353
; Sequence 14353, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

```

; FILE REFERENCE: 1071996.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14353
; LENGTH: 1908
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14353

Query Match      2.9%;   Score 53;   DB 4;   Length 1908;
Best Local Similarity 49.1%;   Pred. No. 0.0042;
Matches 140;   Conservative 0;   Mismatches 145;   Indels 0;   Gaps 0;

Qy 321 CACCATTAGTGTGGACAAGTGGTGAATATCAAGGCCAAGGTGAACGGGCGCTTCAACTC 380
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 381 CAGCATCGAGGTGGGCATCCAGTGGCGCTCGGAGGACCTGTGCTCTGAGAAGCAGTGGAA 440
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 948 CGGCGGGGTCTTCGCCCACTGGTTCGSCACCCCTGGAACTGCTCTCCGACTCGGAGCGGA 1007
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 441 TGTGTCAAGGCCCTTGCCCACTTCTGTGGCCCGCGAGAGATCACCAAGGTGAAGCTGAA 500
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1008 GCGGGTCAAGGGCTTGTCTATCAATCGTCTCCGGCGGCAATCGCCCTGTCTGCAACCCGG 1067
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 501 GCAGATCACGCCCGGGAAGAAGAGAGATGGAGACAGTGTGTGGCGCTCAGCGCCG 560
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1068 CCTCGACTGCTGGAGGACGCACCGCGAAGCCGTGCTCGGGTGCTGCCCTATGTCAG 1127
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 561 GCGCATCGGCCCTGTCTATCGACACCATCAAGGACCTCTGGC 605
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1128 CGACTCTGCATCTGGAGCGGAGGAGCGCATTCGACACGCGCAGGC 1172
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
US-09-328-111-475
; Sequence 475, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 475
; LENGTH: 647
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(647)
; OTHER INFORMATION: n = A.T.C.G

```

US-09-328-111-475

Query Match 2.7%; Score 50.6; DB 3; Length 647;
Best Local Similarity 86.2%; Pred. No. 0.011; 9; Indels 0; Gaps 0;
Matches 56; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1793 TCCCGGTGGCCAAATCAACAGCGGATTTGGATACCTTCAAGACACCTGAAACCTTATCA 1852
DB 289 TTCTCCAGGCCATATCATATAGACCTATTGGATACCATCAAGACACCTGAAACCTTATCG 348
QY 1853 TGAGC 1857
DB 349 TGAGC 353

RESULT 12

US-09-252-991A-5935
; Sequence 5935, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5935
; LENGTH: 975
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5935

Query Match 2.6%; Score 49.2; DB 4; Length 975;
Best Local Similarity 45.5%; Pred. No. 0.027;
Matches 174; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
QY 1176 CTTGAGCTACATACATACGCTCTCTCTTGAAGATGCTTGTGGCCAAAGGCAACTGGGTGCT 1235
DB 67 CCAGGCCACGACGCGCTGCTCAAGGCTTCTCTGAGGAAATCGGTACTCTGCTGCC 126
QY 1236 GTCCTCGGAGATCAGTCAGTCAGTCGCTGTACACTCTGGAGGATGCAAGTTCCTCTCTT 1295
DB 127 CGAGCGGAAGACTTCCAGGCGCGCACCCAGAACTGCGACGAGATGCGCCGATGCG 186
QY 1296 CCACATGAGATGCTGCTGATGCTGATGCGAGCCGCTTCTCTGCTGCTCTCGACCT 1355
DB 187 CGGCGCGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 246
QY 1356 GCGTCAGAGGCGCAGATGGGCAACAGACTACCGAGAGCGTGGAGCTAGTGCACAGGTAGA 1415
DB 247 CGGCTGGGCTCGCTGTACAGGCACTCTACGCGCACCGACGCTGATCAGCGAAGAGGCGG 306
QY 1416 CGAGACGACGCACTTACACGCTCAGCAGCCCTCGCCCTCGAGGCTCACAAAGCCCA 1475
DB 307 CGCCGAGAAAGGCAAGGGCTACAAAGGTTCCGCGGCAAGGTCATCGCTTCGCTCG 366
QY 1476 GGACTTCGTGATCTCTGGCTCGAGGCGGAAGCTTTGTGACAATGGGACCCCTATGTCTAT 1535
DB 367 CGCTCTCTCGAGGCGCGCGCTGGAGTCCGCTCCCATGTGCGAGGCCACTTCTCTA 426
QY 1536 CGCGCTGAGTGGTCAACGCTG 1557
DB 427 CAGCGTGAAGAGCGGCGCTG 448

RESULT 13

US-09-252-991A-5953
; Sequence 5953, Application US/09252991A

Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5953
; LENGTH: 2196
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5953

Query Match 2.6%; Score 49.2; DB 4; Length 2196;
Best Local Similarity 45.5%; Pred. No. 0.033;
Matches 174; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
QY 1176 CTTGAGCTACATACATACGCTCTCTCTTGAAGATGCTTGTGGCCAAAGGCAACTGGGTGCT 1235
DB 237 CCAGGCCACGACGCGCTGCTCAAGGCTTCTCTGAGGAAATCGGTACTCTGCTGCC 296
QY 1236 GTCCTCGGAGATCAGTCAGTCAGTCGCTGTACACTCTGGAGGATGCAAGTTCCTCTCTT 1295
DB 297 CGAGCGGAAGACTTCCAGGCGCGCACCCAGAACTGCGACGAGATGCGCCGATGCG 356
QY 1296 CCACATGAGATGCTGCTGATGCTGATGCTGAGTGCAGCCAGGCTTCTCTGCTGCTCTCGACCT 1355
DB 357 CGGCGCGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 416
QY 1356 GCGTCAGAGGCGCAGATGGGCAACAGACTACCGAGCGTGGAGCTAGTGCACAGGTAGA 1415
DB 417 CGCTGGGCTGCTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 476
QY 1416 CGAGACGACGCACTTACACGCTCAGCAGCTCAGCAGCCCTCGGAGGTACACAAAGCCCA 1475
DB 477 CGCCGAGAAAGGCAAGGCTACAAAGTTCGCGGCGCAAGGTCATCGCTTCGCTCG 536
QY 1476 GGACTTCGTGATCTCTGGCTCGAGGCGGAAGCTTTGTGACAATGGGACCCCTATGTCTAT 1535
DB 537 CGCTTCTCGACGAGGCGCGCTGAGTCCGCTCCCATGTGCGAGCCACTTCTCTA 596
QY 1536 CGCGCTGAGTGGTCAACGCTG 1557
DB 597 CAGCGTGAAGAACGCGCGCTG 618

RESULT 14

US-09-252-991A-5934/G
; Sequence 5934, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5934
; LENGTH: 3036
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5934

Blank Sheet